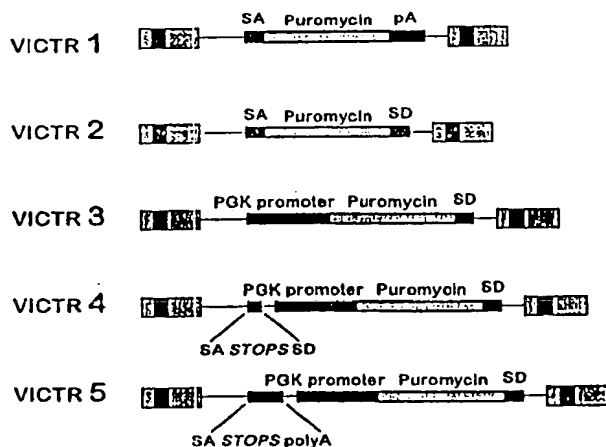




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68, C12N 5/02, 5/06, 15/00, 15/64, C07H 21/04		A1	(11) International Publication Number: WO 98/14614
			(43) International Publication Date: 9 April 1998 (09.04.98)
(21) International Application Number: PCT/US97/17791		(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, HU, ID, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 3 October 1997 (03.10.97)			
(30) Priority Data:			
08/726,867	4 October 1996 (04.10.96)	US	
08/728,963	11 October 1996 (11.10.96)	US	
08/907,598	8 August 1997 (08.08.97)	US	
(71) Applicant: LEXICON GENETICS INCORPORATED [US/US]; 4000 Research Forest Drive, The Woodlands, TX 77381 (US).		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(72) Inventors: SANDS, Arthur; 163 Bristol Bend Circle, The Woodlands, TX 77382 (US). FRIEDRICH, Glenn; 30 Reflection Point, The Woodlands, TX 77381 (US). ZAMBROWICZ, Brian; 18 Firethorne Place, The Woodlands, TX 77382 (US). BRADLEY, Allan; 5127 Queensloch, Houston, TX 77096 (US).			
(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).			

(54) Title: **AN INDEXED LIBRARY OF CELLS CONTAINING GENOMIC MODIFICATIONS AND METHODS OF MAKING AND UTILIZING THE SAME**



(57) Abstract

Methods and vectors (both DNA and retroviral) are provided for the construction of a Library of mutated cells. The Library will preferably contain mutations in essentially all genes present in the genome of the cells. The nature of the Library and the vectors allow for methods of screening for mutations in specific genes, and for gathering nucleotide sequence data from each mutated gene to provide a database of tagged gene sequences. Such a database provides a means to access the individual mutant cell clones contained in the Library. The invention includes the described Library, methods of making the same, and vectors used to construct the Library. Methods are also provided for accessing individual parts of the Library either by sequence or by pooling and screening. The invention also provides for the generation of non-human transgenic animals which are mutant for specific genes as isolated and generated from the cells of the Library.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

AN INDEXED LIBRARY OF CELLS CONTAINING GENOMIC MODIFICATIONS
AND METHODS OF MAKING AND UTILIZING THE SAME

The present application claims priority to U.S.
5 Applications Ser. Nos. 08/726,867, filed October 4, 1996,
08/728,963, filed October 11, 1996, and 08/907,598, filed
August 8, 1997, the disclosures of which are herein
incorporated by reference.

10

1.0. FIELD OF THE INVENTION

The invention relates to an indexed library of
genetically altered cells and methods of organizing the cells
into an easily manipulated and characterized Library. The
invention also relates to methods of making the library,
15 vectors for making insertion mutations in genes, methods of
gathering sequence information from each member clone of the
Library, and methods of isolating a particular clone of
interest from the Library.

20

2.0. BACKGROUND OF THE INVENTION

The general technologies of targeting mutations into the
genome of cells, and the process of generating mouse lines
from genetically altered embryonic stem (ES) cells with
specific genetic lesions are well known (Bradley, 1991, Cur.
25 Opin. Biotech. 2:823-829). A random method of generating
genetic lesions in cells (called gene, or promoter, trapping)
has been developed in parallel with the targeted methods of
genetic mutation (Allen et al., 1988 Nature 333(6176):852-
855; Brenner et al., 1989, Proc. Natl. Acad. Sci. U.S.A.
30 86(14):5517-5521; Chang et al., 1993, Virology 193(2):737-
747; Friedrich and Soriano, 1993, Insertional mutagenesis by
retroviruses and promoter traps in embryonic stem cells, p.
681-701. In Methods Enzymol., vol. 225., P. M. Wassarman and
M. L. DePamphilis (ed.), Academic Press, Inc., San Diego;
35 Friedrich and Soriano, 1991, Genes Dev. 5(9):1513-1523;
Gossler et al., 1989, Science 244(4903):463-465; Kerr et al.,
1989, Cold Spring Harb. Symp. Quant. Biol. 2:767-776; Reddy
et al., 1991, J Virol. 65(3):1507-1515; Reddy et al., 1992,

Proc. Natl. Acad. Sci. U.S.A. 89(15):6721-6725; Skarnes et al., 1992, Genes Dev. 6(6):903-918; von Melchner and Ruley, 1989, J. Virol. 63(8):3227-3233; Yoshida et al., 1995, Transgen. Res. 4:277-287). Gene trapping provides a means to
5 create a collection of random mutations by inserting fragments of DNA into transcribed genes. Insertions into transcribed genes are selected over the background of total insertions since the mutagenic DNA encodes an antibiotic resistance gene or some other selectable marker. The
10 selectable marker lacks its own promoter and enhancer and must be expressed by the endogenous sequences that flank the marker after it has integrated. Using this approach, transcription of the selectable marker is activated and the cell gene is concurrently mutated. This type of strict
15 selection makes it possible to easily isolate thousands of ES cell colonies, each with a unique mutagenic insertion.

Collecting mutants on a large-scale has been a powerful genetic technique commonly used for organisms which are more amenable to such analysis than mammals. These organisms,
20 such as *Drosophila melanogaster*, yeast *Saccharomyces cerevisiae*, and plants such as *Arabidopsis thaliana* are small, have short generation times and small genomes (Bellen et al., 1989, Genes Dev. 3(9):1288-1300; Bier et al., 1989, Genes Dev. 3(9):1273-1287; Hope, 1991, Develop. 113(2):399-408.
25 These features allow an investigator to rear many thousands or millions of different mutant strains without requiring unmanageable resources. However, these type of organisms have only limited value in the study of biology relevant to human physiology and health. It is therefore important to
30 have the power of large-scale genetic analysis available for the study of a mammalian species that can aid in the study of human disease. Given that the entire human genome is presently being sequenced, the comprehensive genetic analysis of a related mammalian species will provide a means to
35 determine the function of genes cloned from the human genome. At present, rodents, and particularly mice, provide the best model for genetic manipulation and analysis of mammalian

physiology.

Gene trapping has been used as an analytical tool to identify genes and regulatory regions in a variety of animal cell types. One system that has proved particularly useful is based on the use of ROSA (reverse orientation splice acceptor) retroviral vectors (Friedrich and Soriano, 1991 and 1993).

The ROSA system can generate mutations that result in a detectable homozygous phenotype with a high frequency. About 50% of all the insertions caused embryonic lethality. The specifically mutated genes may easily be cloned since the gene trapping event produces a fusion transcript. This fusion transcript has trapped exon sequences appended to the sequences of the selectable marker allowing the latter to be used as a tag in polymerase chain reaction (PCR)-based protocols, or by simple cDNA cloning. Examples of genes isolated by these methods include a transcription factor related to human TEF-1 (transcription enhancer factor-1) which is required in the development of the heart (Chen et al., 1994, Genes Devel. 8:2293-2301. Another (spock), is distantly related to yeast genes encoding secretion proteins and is important during gastrulation.

The above experiments have established that the ROSA system is an effective analytical tool for genetic analysis in mammals. However, the structure of many ROSA vectors selects for the "trapping" of 5' exons which, in many cases, do not encode proteins. Such a result is adequate where one wishes to identify and eventually clone control (i.e., promoter or enhancer) sequences, but is not optimal where the generation of insertion-inactivated null mutations is desired, and relevant coding sequence is needed. Thus, the construction of large-scale mutant (preferably null mutant) libraries requires the use of vectors that have been designed to select for insertion events that have occurred within the coding region of the mutated genes as well as vectors that are not limited to detecting insertions into expressed genes.

3.0. SUMMARY OF THE INVENTION

An object of the present invention is to provide a set of genetically altered cells (the 'Library'). The genetic alterations are of sufficient randomness and frequency such
5 that the combined population of cells in the Library represent mutations in essentially every gene found in the cell's genome. The Library is used as a source for obtaining specifically mutated cells, cell lines derived from the individually mutated cells, and cells for use in the
10 production of transgenic non-human animals.

A further object is to provide the vectors, both DNA and retroviral based, that may be used to generate the Library. Typically, at least two distinct vector designs will be used in order to mutate genes that are actively expressed in the
15 target cell, and genes that are not expressed in the target cell. Combining the mutant cells obtained using both types of vectors best ensures that the Library provides a comprehensive set of gene mutations.

A particularly useful vector class contemplated by the
20 present invention includes a vector for inserting foreign exons into animal cell transcripts that comprises a selectable marker, a promoter element operatively positioned 5' to the selectable marker, a splice donor site operatively positioned 3' to the selectable marker, and a second
25 mutagenic foreign polynucleotide sequence located upstream from the promoter element that disrupts, or otherwise "poisons", the splicing or read-through expression of the endogenous cellular transcript. Typically, the mutagenic foreign polynucleotide sequence may incorporate a
30 polyadenylation (pA) site, a nested set of stop codons in each of the three reading frames, splice acceptor and splice donor sequences in operable combination, a mutagenic exon, or any mixture of mutagenic features that effectively prevent the expression of the cellular gene. For example, a
35 polyadenylation sequence may be incorporated in addition to or in lieu of the splice donor sequence. A preferred organization for the mutagenic polynucleotide sequence

comprises a polyadenylation site positioned upstream from a selectable marker which is in turn located upstream from a splice acceptor sequence. Preferably, such a vector does not comprise a transcription terminator or polyadenylation site 5 operatively positioned relative to the coding region of the selectable marker, and shall not comprise a splice acceptor site operatively positioned between the promoter element and the initiation codon of said selectable marker.

An additional vector contemplated by the present 10 invention is designed to replace the normal 3' end of an animal cell transcript with a foreign exon. Such a vector shall generally be engineered to comprise a selectable marker, a splice acceptor site operatively positioned upstream (5') from the initiation codon of the selectable 15 marker, and a polyadenylation site operatively positioned downstream (3') from the termination codon (3' end) of the selectable marker. Preferably, the vector will not comprise a promoter element operatively positioned upstream from the coding region of the selectable marker, and will not comprise 20 a splice donor sequence operatively positioned between the 3' end of the coding region of the selectable marker and the polyadenylation site.

Yet another vector contemplated by the present invention is a vector designed to insert a mutagenic foreign 25 polynucleotide sequence within an animal cell transcript (i.e., the foreign polynucleotide sequence is flanked on both sides by endogenous exons). As described above, the mutagenic foreign polynucleotide sequence may be any sequence that disrupts the normal expression of the gene into which 30 the vector has integrated. Optionally, the vector may additionally incorporate a selectable marker, a splice acceptor site operatively positioned 5' to the initiation codon of the selectable marker, a splice donor site operatively positioned 3' to said selectable marker. 35 Preferably, this vector shall not comprise a polyadenylation site operatively positioned 3' to the coding region of said selectable marker, and shall not comprise a promoter element

operatively positioned 5' to the coding region of said selectable marker.

An additional embodiment of the present invention is a library of genetically altered cells that have been treated 5 to stably incorporate one or more types of the vectors described above. The presently described library of cultured animal cells may be made by a process comprising the steps of treating (i.e., infecting, transfecting, retrotransposing, or virtually any other method of 10 introducing polynucleotides into a cell) a population of cells to stably integrate a vector that mediates the splicing of a foreign exon internal to a cellular transcript, transfecting another population of cells to stably integrate a vector that mediates the splicing of a foreign exon 5' to 15 an exon of a cellular transcript, and selecting for transduced cells that express the products encoded by the foreign exons.

Alternatively, an additional embodiment of the present invention describes a mammalian cell library made by a method 20 comprising the steps of: transfecting a population of cells with a vector capable of expressing a selectable marker in the cell only after the vector inserts into the host genome; transfecting or infecting a population of cells with a vector containing a selectable marker that is substantially only 25 expressed by cellular control sequences (after the vector integrates into the host cells genome); and growing the transfected cells under conditions that select for the expression of the selectable marker.

In an additional embodiment of the present invention, 30 the two populations of transfected cells will be individually grown under selective conditions, and the resulting mutated population of cells collectively comprises a substantially comprehensive library of mutated cells.

In an additional embodiment of the present invention, 35 the individual mutant cells in the library are separated and clonally expanded. Additionally, the clonally expanded mutant cells may then be analyzed to ascertain the DNA

sequence, or partial DNA sequence of the mutated host gene.

The presently described methods of making, organizing, and indexing libraries of mutated animal cells are also broadly applicable to virtually any eukaryotic cells that may
5 be genetically manipulated and grown in culture.

The invention provides for sequencing every gene mutated in the Library. The resulting sequence database subsequently serves as an index for the library. In essence, every cell line in the Library is individually catalogued using the
10 partial sequence information. The resulting sequence is specific for the mutated gene since the present methods are designed to obtain sequence information from exons that have been spliced to the marker sequence. Since the coverage of the mutagenesis is preferably the entire set of genes in the
15 genome, the resulting Library sequence database contains sequence from essentially every gene in the cell. From this database, a gene of interest can be identified. Once identified, the corresponding mutant cell may be withdrawn from the Library based on cross reference to the sequence
20 data.

An additional embodiment of the invention provides for methods of isolating mutations of interest from the Library. Two methods are proposed for obtaining individual mutant cell lines from the Library. The first provides a scheme where
25 clones of the cells generated using the above vectors are pooled into sets of defined size. Using the procedure described below which utilizes reverse transcription (RT) and polymerase chain reaction (PCR), a cell line with a mutation in a gene whose sequence is partly or wholly known is
30 isolated from organized sets of these pools. A few rounds of this screening procedure results in the isolation of the desired individual cell line.

A second procedure involves the sequencing of regions flanking the vector insertion sites in the various cells in
35 the library. The sequence database generated from these data effectively constitutes an index of the clones in the library that may be used to identify cells having mutations in

specific genes.

4.0. DESCRIPTION OF THE FIGURES

Figure 1. Shows a diagrammatic representation of 5 different
5 vectors that are generally representative of the type of
vectors that may be used in the present invention.

Figure 2. Shows a general strategy for identifying "trapped"
cellular sequences by PCR analysis of the cellular exons that
10 flank the foreign intron introduced by the VICTR 2 vector.

Figure 3 shows a PCR based strategy for identifying tagged
genes by chromosomal location.

15 Figure 4. Is a diagrammatic representation of a strategy of
identifying or indexing the specific clones in the library
via PCR analysis and sequencing of mRNA samples obtained from
the cells in the library.

20 Figure 5. Is a diagrammatic representation of a method of
isolating positive clones by screening pooled mutant cell
clones.

Figure 6. Partial nucleic acid or predicted amino acid
25 sequence data from 9 clones (OST1-9) isolated using the
described techniques aligned with similar sequences from
previously characterized genes.

Figure 7. Provides a diagrammatic representation of VICTRs 3
30 and 20 as well as the transcripts that result after
integration into a hypothetical region of the target cell
genome (i.e., "Wildtype Locus").

Figure 8. Provides a representative list of a portion of the
35 known genes that have been identified using the disclosed
methods and technology.

5.0. DETAILED DESCRIPTION OF THE INVENTION

The present invention describes a novel indexed library containing a substantially comprehensive set of mutations in the host cell genome, and methods of making and using the same. The presently described Library comprises as a set of cell clones that each possess at least one mutation (and preferably a single mutation) caused by the insertion of DNA that is foreign to the cell. For the purposes of the present invention, "foreign" polynucleotide sequences can be any sequences that are newly introduced to a cell, do not naturally occur in the cell at the engineered region of the chromosome, or occur in the cell but are not organized to provide an identical function to that provided in the engineered vector.

The particularly novel features of the Library include the methods of construction, and indexing. To index the library, the mutant cells of the library are clonally expanded and each mutated gene is at least partially sequenced. The Library thus provides a novel tool for assessing the specific function of a given gene. The insertions cause a mutation which allow for essentially every gene represented in the Library to be studied using genetic techniques either *in vitro* or *in vivo* (via the generation of transgenic animals). For the purposes of the present invention, the term "essentially every gene" shall refer to the statistical situation where there is generally at least about a 70 percent probability that the genomes of cells used to construct the library collectively contain at least one inserted vector sequence in each gene, preferably a 85 percent probability, and more specifically at least about a 95 percent probability as determined by a standard Poisson distribution.

Also for the purposes of the present invention the term "gene" shall refer to any and all discrete coding regions of the cell's genome, as well as associated noncoding and regulatory regions. Additionally, the term operatively positioned shall refer to the control elements or genes that

are provided with the proper orientation and spacing to provide the desired or indicated functions of the control elements or genes.

For the purposes of the present invention, a gene is "expressed" when a control element in the cell mediates the production of functional or detectable levels of mRNA encoded by the gene, or a selectable marker inserted therein. A gene is not expressed where the control element in the cell is absent, has been inactivated, or does not mediate the production of functional or detectable levels of mRNA encoded by the gene, or a selectable marker inserted therein.

5.1. Vectors used to build the Library

A number of investigators have developed gene trapping vectors and procedures for use in mouse and other cells (Allen et al., 1988; Bellen et al., 1989, Genes Dev. 3(9):1288-1300; Bier et al., 1989, Genes Dev. 3(9):1273-1287; Bonnerot et al., 1992, J Virol. 66(8):4982-4991; Brenner et al., 1989; Chang et al., 1993; Friedrich and Soriano, 1993; Friedrich and Soriano, 1991; Goff, 1987, Methods Enzymol. 152:469-481; Gossler et al.; Hope, 1991; Kerr et al., 1989; Reddy et al., 1991; Reddy et al., 1992; Skarnes et al., 1992; von Melchner and Ruley; Yoshida et al., 1995). The gene trapping system described in the present invention is based on significant improvements to the published SA (splice acceptor) DNA vectors and the ROSA (reverse orientation, splice acceptor) retroviral vectors (Chen et al., 1994; Friedrich and Soriano, 1991 and 1993). The presently described vectors also use a selectable marker called β geo. This gene encodes a protein which is a fusion between the β -galactosidase and neomycin phosphotransferase proteins. The presently described vectors place a splice acceptor sequence upstream from the β geo gene and a poly-adenylation signal sequence downstream from the marker. The marker is integrated after transfection by, for example, electroporation (DNA vectors), or retroviral infection, and gene trap events are selected based on resistance to G418

resulting from activation of β geo expression by splicing from the endogenous gene into the ROSA splice acceptor. This type of integration disrupts the transcription unit and preferably results in a null mutation at the locus.

- 5 Although gene trapping has proven a useful analytical tool, the present invention contemplates gene trapping on a large scale. The vectors utilized in the present invention have been engineered to overcome the shortcomings of the early gene trap vector designs, and to facilitate procedures
- 10 allowing high throughput. In addition, procedures are described that allow the rapid and facile acquisition of sequence information from each trapped cDNA which may be adapted to allow complete automation. These latter procedures are also designed for flexibility so that
- 15 additional molecular information can easily be obtained subsequently. The present invention therefore incorporates gene trapping into a larger and unique tool. A specially organized set of gene trap clones that provide a novel and powerful new tool of genetic analysis.
- 20 The presently described vectors are superficially similar to the ROSA family of vectors, but constitute significant improvements and provide for additional features that are useful in the construction and indexing of the Library. Typically, gene trapping vectors are designed to
- 25 detect insertions into transcribed gene regions within the genome. They generally consist of a selectable marker whose normal expression is handicapped by exclusion of some element required for proper transcription. When the vector integrates into the genome, and acquires the necessary
- 30 element by juxtaposition, expression of the selectable marker is activated. When such activation occurs, the cell can survive when grown in the appropriate selective medium which allows for the subsequent isolation and characterization of the trapped gene. Integration of the gene trap generally
- 35 causes the gene at the site of integration to be mutated.

Some gene trapping vectors have a splice acceptor preceding a selectable marker and a poly-adenylation signal

following the selectable marker, and the selectable marker gene has its own initiator ATG codon. Using this arrangement, the fusion transcripts produced after integration generally only comprise exons 5' to the insertion site to the known marker sequences. Where the vector has inserted into the 5' region of the gene, it is often the case that the only exon 5' to the vector is a non-coding exon. Accordingly, the sequences obtained from such fusions do not provide the desired sequence information about the relevant gene products. This is because untranslated sequences are generally less well conserved than coding sequences.

To compensate for the short-comings of earlier vectors, the vectors of the present invention have been designed so that 3' exons are appended to the fusion transcript by replacing the poly-adenylation and transcription termination signals of earlier ROSA vectors with a splice donor (SD) sequence. Consequently transcription and splicing generally results in a fusion between all or most of the endogenous transcript and the selectable marker exon, for example β geo, neomycin (neo) or puromycin (puro). The exon sequences immediately 3' to the selectable marker exon may then be sequenced and used to establish a database of expressed sequence tags. The presently described procedures will typically provide approximately 200 nucleotides of sequence, or more. These sequences will generally be coding and therefore informative. The prediction that the sequence obtained will be from coding region is based on two factors. First, gene trap vectors are generally found near the 5' end of the gene immediately after untranslated exons because the method selects for integration events that place the initiator ATG of the selectable marker as the first encountered, and thus used, for translation. Second, mammalian transcripts have short 5' untranslated regions (UTRs) which are typically between 50 and 150 nucleotides in length.

The obtained sequence information also provides a ready source of probes that may be used to isolate the full-length

gene or cDNA from the host cell, or as heterologous probes for the isolation of homologous genes in other species.

Internal exons in mammalian transcripts are generally quite small, on the average 137 bases with few over 300 bases. Consequently, a large internal exon may be spliced less efficiently. Thus, the presently described vectors have been designed to sandwich relatively small selectable markers (for example: *neo*, ~800 bases, or a smaller drug resistance gene such as *puro*, ~600 bases) between the requisite splicing elements to produce relatively small exons. Exons of this size are more typical of mammalian exons and do not present undue problems for the splicing machinery of the cell. Such a design consideration is novel to the presently disclosed gene trapping vectors. Accordingly, an additional embodiment of the claimed vectors is that the respective splice acceptor and splice donor sites are engineered such that they are operatively positioned close to the ends of the selectable marker coding region (the region spanning from the initiation codon to the termination codon). Generally, the splice acceptor or splice donor sequences shall appear within about 80 bases from the nearest end of the selectable marker coding region, preferably within about 50 bases from the nearest end of the coding region, more preferably within about 30 bases from the nearest end of the coding regions and specifically within about 20 bases of the nearest end of the selectable marker coding region.

The new vectors are represented in retroviral form in Figure 1. They are used by infecting target cells with retroviral particles such that the proviruses shown in the schematic can be found in the genome of the target. These vectors are called VICTR which is an acronym for "viral constructs for trapping".

The presently described retroviral vectors may be used in conjunction with retroviral packaging cell lines such as those described in U.S. Patent No. 5,449,614 ("'614 patent") issued September 12, 1995, herein incorporated by reference. Where non-mouse animal cells are to be used as targets for

generating the described libraries, packaging cells producing retrovirus with amphotropic envelopes will generally be employed to allow infection of the host cells.

The mutagenic gene trap DNA may also be introduced into the target cell genome by various transfection techniques which are familiar to those skilled in the art such as electroporation, lipofection, calcium phosphate precipitation, infection, retrotransposition, and the like. Examples of such techniques may be found in Sambrook et al. (1989) Molecular Cloning Vols. I-III, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Current Protocols in Molecular Biology (1989) John Wiley & Sons, all Vols. and periodic updates thereof, herein incorporated by reference. The transfected versions of the retroviral vectors are typically plasmid DNA molecules containing DNA cassettes comprising the described features between the retroviral LTRs.

The vectors VICTR 1 and 2 (Fig. 1) are designed to trap genes that are transcribed in the target cell. To trap genes that are not expressed in the target cell, gene trap vectors such as VICTR 3, 4 and 5 (described below) are provided. These vectors have been engineered to contain a promoter element capable of initiating transcription in virtually any cell type which is used to transcribe the coding sequence of the selectable marker. However, in order to get proper translation of the marker product, and thus render the cell resistant to the selective antibiotic, a polyadenylation signal and a transcription termination sequence must be provided. Vectors VICTR 3 through 5 are constructed such that an effective polyadenylation signal can only be provided by splicing with an externally provided downstream exon that contains a poly-adenylation site. Therefore, since the selectable marker coding region ends only in a splice donor sequence, these vectors must be integrated into a gene in order to be properly expressed. In essence, these vectors append the foreign exon encoding the marker to the 5' end of an endogenous transcript. These events will tag genes and

create mutations that are used to make clones that will become part of the Library.

With the above design considerations, the VICTR series of vectors, or similarly designed and constructed vectors, have the following features. VICTR 1 is a terminal exon gene trap. VICTR 1 does not contain a control region that effectively mediates the expression of the selectable marker gene. Instead, the coding region of the selectable marker contained in VICTR 1, in this case encoding puromycin resistance (but which can be any selectable marker functional in the target cell type), is preceded by a splice acceptor sequence and followed by a polyadenylation addition signal sequence. The coding region of the puro gene has an initiator ATG which is downstream and adjacent to a region of sequence that is most favorable for translation initiation in eukaryotic cells - the so called Kozak consensus sequence (Kozak, 1989, J. Cell, Biol. 108(2):229-241). With a Kozak sequence and an initiator ATG, the puro gene in VICTR 1 is activated by integrating into the intron of an active gene, and the resulting fusion transcript is translated beginning at the puromycin initiation (ATG/AUG) codon. However, terminal gene trap vectors need not incorporate an initiator ATG codon. In such cases, the gene trap event requires splicing and the translation of a fusion protein that is functional for the selectable marker activity. The inserted puromycin coding sequence must therefore be translated in the same frame as the "trapped" gene.

The splice acceptor sequence used in VICTR 1 and other members of the VICTR series is derived from the adenovirus major late transcript splice site located at the intron 1/exon 2 boundary. This sequence contains a polypyrimidine stretch preceding the AG dinucleotide which denotes the actual splice site. The presently described vectors contemplate the use of any similarly derived splice acceptor sequence. Preferably, the splice acceptor site will only rarely, if ever, be involved in alternative splicing events.

The polyadenylation signal at the end of the puro gene is derived from the bovine growth hormone gene. Any similarly derived polyadenylation signal sequence could be used if it contains the canonical AATAAA and can be demonstrated to terminate transcription and cause a polyadenylate tail to be added to the engineered coding exons.

VICTR 2 is a modification of VICTR 1 in which the polyadenylation signal sequence is removed and replaced by a splice donor sequence. Like VICTR 1, VICTR 2 does not contain a control region that effectively mediates the expression of the selectable marker gene. Typically, the splice donor sequence to be employed in a VICTR series vector shall be determined by reference to established literature or by experimentation to identify which sequences properly initiate splicing at the 5' end of introns in the desired target cell. The specifically exemplified sequence, AGGTAAGT, results in splicing occurring in between the two G bases. Genes trapped by VICTR 2 splice upstream exons onto the puro exon and downstream exons onto the end of the puro exon. Accordingly, VICTR 2 effectively mutates gene expression by inserting a foreign exon in-between two naturally occurring exons in a given transcript. Again, the puro gene may or may not contain a consensus Kozak translation initiation sequence and properly positioned ATG initiation codon. As discussed above, gene trapping by VICTR 1 and VICTR 2 requires that the mutated gene is expressed in the target cell line. By incorporating a splice donor into the VICTR traps, transcript sequences downstream from the gene trap insertion can be determined. As described above, these sequences are generally more informative about the gene mutated since they are more likely to be coding sequences. This sequence information is gathered according to the procedures described below.

VICTR 3, VICTR 4 and VICTR 5 are gene trap vectors that do not require the cellular expression of the endogenous trapped gene. The VICTR vectors 3 through 5 all comprise a

promoter element that ensures that transcription of the selectable marker would be found in all cells that have taken up the gene trap DNA. This transcription initiates from a promoter, in this case the promoter element from the mouse phosphoglycerate kinase (PGK) gene. However, since the constructs lack a polyadenylation signal there can be no proper processing of the transcript and therefore no translation. The only means to translate the selectable marker and get a resistant cell clone is by acquiring a polyadenylation signal. Since polyadenylation is known to be concomitant with splicing, a splice donor is provided at the end of the selectable marker. Therefore, the only positive gene trap events using VICTR 3 through 5 will be those that integrate into a gene's intron such that the marker exon is spliced to downstream exons that are properly polyadenylated. Thus genes mutated with the VICTR vectors 3 through 5 need not be expressed in the target cell, and these gene trap vectors can mutate all genes having at least one intron. The design of VICTR vectors 3 through 5 requires a promoter element that will be active in the target cell type, a selectable marker and a splice donor sequence. Although a specific promoter was used in the specific embodiments, it should be understood that appropriate promoters may be selected that are known to be active in a given cell type. Typically, the considerations for selecting the splice donor sequence are identical to those discussed for VICTR 2, *supra*.

VICTR 4 differs from VICTR 3 only by the addition of a small exon upstream from the promoter element of VICTR 4. This exon is intended to stop normal splicing of the mutated gene. It is possible that insertion of VICTR 3 into an intron might not be mutagenic if the gene can still splice between exons, bypassing the gene trap insertion. The exon in VICTR 4 is constructed from the adenovirus splice acceptor described above and the synthetic splice donor also described above. Stop codons are placed in all three reading frames in the exon, which is about 100 bases long. The stops would truncate the endogenous protein and presumably cause a

mutation.

A conceptually similar alternative design uses a terminal exon like that engineered into VICTR 5. Instead of a splice donor, a polyadenylation site is used to terminate transcription and produce a truncated message. Stops in all three frames are also provided to truncate the endogenous protein as well as the resulting transcript.

VICTR 20 is a modified version of VICTR 3 that incorporates a polyadenylation site 5' to the PGK promoter, the IRES β geo sequence (i.e., foreign mutagenic polynucleotide sequence) 5' to the polyadenylation site, and a splice acceptor site 5' to the IRES β geo coding region. VICTR 20 additionally incorporates, in operable combination, a pair of recombinase recognition sites that flank the PGKpuroSD cassette.

All of the traps of the VICTR series are designed such that a fusion transcript is formed with the trapped gene. For all but VICTR 1, the fusion contains cellular exons that are located 3' to the gene trap insertion. All of the flanking exons may be sequenced according to the methods described in the following section. To facilitate sequencing, specific sequences are engineered onto the ends of the selectable marker (e.g., puromycin coding region). Examples of such sequences include, but are not limited to unique sequences for priming PCR, and sequences complementary to the standard M13 forward sequencing primer. Additionally, stop codons are added in all three reading frames to ensure that no anomalous fusion proteins are produced. All of the unique 3' primer sequences are followed immediately by the synthetic 9 base pair splice donor sequence. This keeps the size of the exon comprising the selectable marker (puro gene) at a minimum to best ensure proper splicing, and positions the amplification and sequencing primers immediately adjacent to the flanking "trapped" exons to be sequenced as part of the construction of a Library database.

When any members of the VICTR series are constructed as retroviruses, the direction of transcription of the

selectable marker is opposite to that of the direction of the normal transcription of the retrovirus. The reason for this organization is that the transcription elements such as the polyadenylation signal, the splice sites and the promoter elements found in the various members of the VICTR series interfere with the proper transcription of the retroviral genome in the packaging cell line. This would eliminate or significantly reduce retroviral titers. The LTRs used in the construction of the packaging cell line are self-inactivating. That is, the enhancer element is removed from the 3' U3 sequences such that the proviruses resulting from infection would not have an enhancer in either LTR. An enhancer in the provirus may otherwise affect transcription of the mutated gene or nearby genes.

15 Since a 'cryptic' splice donor sequence is found in the inverted LTRs, this splice donor sequence has been removed from the VICTR vectors by site specific mutagenesis. It was deemed necessary to remove this splice donor so that it would not affect the trapping splicing events.

20 The present disclosure also describes vectors that incorporate a new way to conduct positive selection. VICTR 3 and VICTR 20 are two examples of such vectors. Both VICTR 3 and VICTR 20, contain PGKpuroSD which must splice into exons of gene that provide a polyadenylation addition sequence in order to allow expression of the puromycin selectable marker gene. When placed in a targeting vector, PGKpuroSD allows for positive selection when targeting takes place. In addition to providing positive selection, targeted events among resistant colonies are easy to identify by the 3' RACE protocols (see section 5.2.2., *infra*) used for Omnibank production. This automated process allows for the rapid identification of targeted events. It is important that unlike SA β geo, PGKpuroSD does not require expression of the targeted gene in order to provide positive selection. In addition, VICTR 20 provides 2 potential positive selectable markers (puro and neo). The use of two selectable markers, when a gene is expressed, provides a means to increase the

targeting efficiency by requiring both selectable markers to function which is much more remote a possibility than having one selectable marker function unless there is a targeted event. The addition of a negative selection cassette to
5 these vectors would only increase their targeting efficiency.

An additional feature that may be incorporated into the presently described vectors includes the use of recombinase recognition sequences. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two examples of
10 site-specific DNA recombinase enzymes which cleave DNA at specific target sites (loxP sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. When a piece of DNA is flanked by 2 loxP or frt sites (e.g., recombinase control elements)
15 in the same orientation, the corresponding recombinase will cause the removal of the intervening DNA sequence. When a piece of DNA is flanked by loxP or frt sites in an indirect orientation, the corresponding recombinase will essentially activate the control elements to cause the intervening DNA to
20 be flipped into the opposite orientation. These recombinases provide powerful approaches for manipulating DNA *in situ*.

Recombinases have important applications for gene trapping and the production of a library of trapped genes. When constructs containing PGKpuroSD are used to trap genes,
25 the fusion transcript between puromycin and sequences of the trapped gene could result in some level of protein expression from the trapped gene if translational reinitiation occurs. Another important issue is that several reports suggest that the PGK promoter can affect the expression of nearby genes.
30 These effects may make it difficult to determine gene function after a gene trap event since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase
35 activity. When PGKpuroSD is flanked by loxP, frt, or any other recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal

of PGKpuroSD. In this way, effects caused by PGKpuroSD fusion transcripts, or the PGK promoter, are avoided.

Accordingly, a vector that may be particularly useful for the practice of the present invention is VICTR 20. This 5 vector replaces the terminal exon of VICTR 5 with a splice acceptor located upstream from the β geo gene which can be used for both LacZ staining and antibiotic selection. The fusion gene possesses its own initiator methionine and an internal ribosomal entry site (IRES) for efficient 10 translation initiation. In addition, the PGK promoter and puromycin-splice donor sequences have been flanked by lox P recombination sites. This allows for the ability to both remove and introduce sequences at the integration site and is of potential value with regard to the manipulation of regions 15 proximal to trapped target genes (Barinaga, Science 265:26-8, 1994). While this particular vector includes lox P recombination sites, the present invention is in no way limited to the use of this specific recombination site (Akagi et al., Nucleic Acids Res 25:1766-73, 1997).

20 Another very important use of recombinases is to produce mutations that can be made tissue-specific and/or inducible. In the presently described vectors, the SA β geo or SAIRES β geo component provides the mutagenic function by "trapping" the normal splicing from preceding exons. If the SA β geo is 25 flanked by inverted loxP, frt, or any other recombinase sites, the addition of the corresponding recombinase results in the flipping of the SA β geo sequence so that it no longer prevents the normal splicing of the cellular gene into which it is integrated. To make a gene trap tissue-specific or 30 inducible one could produce the trap with SA β geo in the reverse orientation and then provide recombinase activity only at the time and place where one wishes to remove the gene function. The use of tissue-specific or inducible recombinase constructs allows one to choose when and where 35 one removes, or activates, the function of the targeted gene.

One method for practicing the inducible forms of recombinase mediated gene expression involves the use of

vectors that use inducible or tissue specific promoter/operator elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or
5 activation of expression of the desired recombinase activity. Examples of such inducible promoters or control elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No
10 et al., Proc Natl Acad Sci USA 93:3345-51, 1996; Furth et al., Proc Natl Acad Sci USA 91:9302-6, 1994). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. Vectors incorporating such promoters would
15 only express recombinase activity in cells that express the necessary transcription factors.

The incorporation of recombinase sites into the gene trapping vectors highlights the value of using the described gene trap vectors to deliver specific DNA sequence elements
20 throughout the genome. Although a variety of vectors are available for placing sequences into the genome, the presently described vectors facilitate both the insertion of the specific elements, and the subsequent identification of where sequence has inserted into the cellular chromosome.
25 Additionally, the presently described vectors may be used to place recombinase recognition sites throughout the genome. The recombinase recognition sites could then be used to either remove or insert specific DNA sequences at predetermined locations.

30 Moreover, the described gene trap vectors can also be used to insert regulatory elements throughout the genome. Recent work has identified a number of inducible or repressible systems that function in the mouse. These include the rapamycin, tetracycline, ecdysone,
35 glucocorticoid, and heavy metal inducible systems. These systems typically rely on placing DNA elements in or near a promoter. An inducible or repressible transcription factor

that can identify and bind to the DNA element may also be engineered into the cells. The transcription factor will specifically bind to the DNA element in either the presence or absence of a ligand that binds to the transcription factor and, depending on the structure of the transcription factor, it will either induce or repress the expression of the cellular gene into which the DNA elements have been inserted. The ability to place these inducible or repressible elements throughout the genome would increase the value of the library by adding the potential to regulate the expression of the trapped gene.

The vectors described also have important applications for the overexpression of genes or portions of genes to select for phenotypic effects. Currently, overexpression of cDNA libraries to look for genes or parts of genes with specific functions is a common practice. One example would be to overexpress genes or portions of genes to look for expression that causes loss of contact inhibition for cell growth as determined by growth in soft agar. This would allow the identification of genes or portions of genes that can act as oncogenes. Simple modifications of VICTR 20 would allow it to be used for these applications. For example, the addition of an internal ribosome entry site (IRES) 3' to the puromycin selectable marker and before the SD sequence, would result in the overexpression of sequences from the trapped downstream exons. In addition, the IRES could be modified by, for example, the addition of one or two nucleotides such that there could be 3 basic vectors that would allow expression of trapped exons in all three reading frames. In this way, genes could be trapped throughout the genome resulting in overexpression of genes, or portions thereof, to examine the cellular function of the trapped genes. This identification of function could be done by selecting for the function of interest (i.e., growth in soft agar could result from the overexpression of potentially oncogenic genes). This technique would allow for the screening or selection of large numbers of genes, or portions thereof, by

overexpressing the genes and identifying cells displaying the phenotypes of interest. Additional assays could, for example, identify candidate tumor suppressor genes based on their ability, when overexpressed, to prevent growth in soft agar.

Given the fact that expression pattern information can provide insight into the possible functions of genes mutated by the current methods, another LTR vector, VICTR 6, has been constructed in a manner similar to VICTR 5 except that the terminal exon has been replaced with either a gene coding for β -galactosidase (β gal) or a fusion between β -gal and neomycin phosphotransferase (β geo), each preceded by a splice acceptor and followed by a polyadenylation signal. Endogenous gene expression and splicing of these markers into cellular transcripts and translation into fusion proteins will allow for increased mutagenicity as well as the delineation of expression through Lac Z staining.

An additional vector, VICTR 12, incorporates two separate selectable markers for the analysis of both integration sites and trapped genes. One selectable marker (e.g. puro) is similar to that for VICTRs 3 through 5 in that it contains a promoter element at its 5' end and a splice donor sequence 3'. This gene cassette is located in the LTRs of the retroviral vector. The other marker (neo) also contains a promoter element but has a polyadenylation signal present at the 3' end of the coding sequence and is positioned between the viral LTRs. Both selectable markers contain an initiator ATG for proper translation. The design of VICTR 12 allows for the assessment of absolute titer as assayed by the number of colonies resistant to antibiotic selection for the constitutively expressed marker possessing a polyadenylation signal. This titer can then be compared to that observed for gene-trapping and stable expression of the resistance marker flanked at its 3' end by a splice donor. These numbers are important for the calculation of gene trapping frequency in the context of both nonspecific binding by retroviral integrase and directed binding by chimeric

integrase fusions. In addition, it provides an option to focus on the actual integration sites through infection and selection for the marker containing the polyadenylation signal. This eliminates the need for the fusion protein binding to occur upstream and in the proximity of the target gene. Theoretically, any transcription factor binding sites present within the genome are targets for proximal integration and subsequent antibiotic resistance. Analysis of sequences flanking the LTRs of the retroviral vector should reveal canonical factor binding sites. In addition, by including the promoter/splice donor design of VICTR 3, gene-trapping abilities are retained in VICTR 12.

VICTR A is a vector which does not contain gene trapping constructs but rather a selectable marker possessing all of the required entities for constitutive expression including, but not limited to, a promoter element capable of driving expression in eukaryotic cells and a polyadenylation and transcriptional terminal signal. Similar to VICTR 12, downstream gene trapping is not necessary for successful selection using VICTR A. This vector is intended solely to select for successful integrations and serves as a control for the identification of transcription factor binding sites flanking the integrant as mentioned above.

Finally, VICTR B is similar to VICTR A in that it comprises a constitutively expressed selectable marker, but it also contains the bacterial β -lactamase ampicillin resistance selectable marker and a ColE1 origin of replication. These entities allow for the rapid cloning of sequences flanking the long terminal repeats through restriction digestion of genomic DNA from infected cells and ligation to form plasmid molecules which can be rescued by bacterial transformation, and subsequently sequenced. This vector allows for the rapid analysis of cellular sequences that contain putative binding sites for the transcription factor of interest.

Other vector designs contemplated by the present invention are engineered to include an inducible regulatory

elements such as tetracycline, ecdysone, and other steroid-responsive promoters (No et al., Proc Natl Acad Sci USA 93:3345-51, 1996; Furth et al., Proc Natl Acad Sci USA 91:9302-6, 1994). These elements are operatively positioned to allow the inducible control of expression of either the selectable marker or endogenous genes proximal to site of integration. Such inducibility provides a unique tool for the regulation of target gene expression.

All of the gene trap vectors of the VICTR series, with the exception of VICTRs A and B, are designed to form a fusion transcript between vector encoded sequence and the trapped target gene. All of the flanking exons may be sequenced according to the methods described in the following section. To facilitate sequencing, specific sequences are engineered onto the ends of the selectable marker (e.g., puromycin coding region). Examples of such sequences include, but are not limited to unique sequences for priming PCR, and sequences complementary to standard M13 sequencing primers. Additionally, stop codons are added in all three reading frames to ensure that no anomalous fusion proteins are produced. All of the unique 3' primer sequences are immediately followed by a synthetic 9 base pair splice donor sequence. This keeps the size of the exon comprising the selectable marker at a minimum to ensure proper splicing, and positions the amplification and sequencing primers immediately adjacent to the flanking trapped exons to be sequenced as part of the generation of the collection of cells representing mutated transcription factor targets.

Since a cryptic splice donor sequence is found in the inverted LTRs, this cryptic splice donor sequence has been removed from the VICTR vectors by site specific mutagenesis. It was deemed necessary to remove this splice donor so that it would not affect trapping associated splicing events.

When any members of the VICTR series are packaged into infectious virus, the direction of transcription of the selectable marker is opposite to that of the direction of the normal transcription of the retrovirus. The reason for this

organization is that the regulatory elements such as the polyadenylation signal, the splice sites and the promoter elements found in the various members of the VICTR series can interfere with the transcription of the retroviral genome in the packaging cell line. This potential interference may significantly reduce retroviral titers.

Although specific gene trapping vectors have been discussed at length above, the invention is by no means to be limited to such vectors. Several other types of vectors that may also be used to incorporate relatively small engineered exons into a target cell transcripts include, but are not limited to, adenoviral vectors, adenoassociated virus vectors, SV40 based vectors, and papilloma virus vectors. Additionally, DNA vectors may be directly transferred into the target cells using any of a variety of biochemical or physical means such as lipofection, chemical transfection, retrotransposition, electroporation, and the like.

Although, the use of specific selectable markers has been disclosed and discussed herein, the present invention is in no way limited to the specifically disclosed markers. Additional markers (and associated antibiotics) that are suitable for either positive or negative selection of eukaryotic cells are disclosed, *inter alia*, in Sambrook et al. (1989) Molecular Cloning Vols. I-III, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Current Protocols in Molecular Biology (1989) John Wiley & Sons, all Vols. and periodic updates thereof, as well as Table I of U.S. Patent No. 5,464,764 issued November 7, 1995, the entirety of which is herein incorporated by reference. Any of the disclosed markers, as well as others known in the art, may be used to practice the present invention.

5.2. The Analysis of Mutated Genes and Transcripts

The presently described invention allows for large-scale genetic analysis of the genomes of any organism for which there exists cultured cell lines. The Library may be constructed from any type of cell that can be transfected by

standard techniques or infected with recombinant retroviral vectors.

Where mouse ES cells are used, then the Library becomes a genetic tool able to completely represent mutations in
5 essentially every gene of the mouse genome. Since ES cells can be injected back into a blastocyst and become incorporated into normal development and ultimately the germ line, the cells of the Library effectively represent a complete panel of mutant transgenic mouse strains (see
10 generally, U.S. Patent No. 5,464,764 issued November 7, 1995, herein incorporated by reference).

A similar methodology may be used to construct virtually any non-human transgenic animal (or animal capable of being rendered transgenic). Such nonhuman transgenic animals may
15 include, for example, transgenic pigs, transgenic rats, transgenic rabbits, transgenic cattle, transgenic goats, and other transgenic animal species, particularly mammalian species, known in the art. Additionally, bovine, ovine, and porcine species, other members of the rodent family, e.g.
20 rat, as well as rabbit and guinea pig and non-human primates, such as chimpanzee, may be used to practice the present invention.

Transgenic animals produced using the presently described library and/or vectors are useful for the study of
25 basic biological processes and diseases including, but not limited to, aging, cancer, autoimmune disease, immune disorders, alopecia, glandular disorders, inflammatory disorders, diabetes, arthritis, high blood pressure, atherosclerosis, cardiovascular disease, pulmonary disease,
30 degenerative diseases of the neural or skeletal systems, Alzheimer's disease, Parkinson's disease, asthma, developmental disorders or abnormalities, infertility, epithelial ulcerations, and microbial pathogenesis (a relatively comprehensive review of such pathogens is
35 provided, *inter alia*, in Mandell et al., 1990, "Principles and Practice of Infectious Disease" 3rd. ed., Churchill Livingstone Inc., New York, N.Y. 10036, herein incorporated

by reference). As such, the described animals and cells are particularly useful for the practice of functional genomics.

5.2.1. Constructing a Library of Individually Mutated Cell Clones

5 The vectors described in the previous section were used to infect (or transfect) cells in culture, for example, mouse embryonic stem (ES) cells. Gene trap insertions were initially identified by antibiotic resistance (e.g., puromycin). Individual clones (colonies) were moved
10 from a culture dish to individual wells of a multi-welled tissue culture plate (e.g. one with 96 wells). From this platform, the clones were be duplicated for storage and subsequent analysis. Each multi-well plate of clones was
15 then processed by molecular biological techniques described in the following section in order to derive sequence of the gene that has been mutated. This entire process is presented schematically in Figure 4 (described below).

20 5.2.2. Identifying and Sequencing the Tagged Genes in the Library.

The relevant nucleic acid (and derived amino acid sequence information) will largely be obtained using PCR-based techniques that rely on knowing part of the sequence of the fusion transcripts (see generally, Frohman et
25 al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85(23):8998-9000, and U.S. Patents Nos. 4,683,195 to Saiki et al., and 4,683,202 to Mullis, which are herein incorporated by reference). Typically, such sequences are encoded by the foreign exon containing the selectable marker. The procedure
30 is represented schematically in Figure 2 (3' RACE). Although each step of the procedure may be done manually, the procedure is also designed to be carried out using robots that can deliver reagents to multi well culture plates (e.g., but not limited to, 96-well plates).

35 The first step generates single stranded complementary DNA which is used in the PCR amplification reaction (Figure

2). The RNA substrate for cDNA synthesis may either be total cellular RNA or an mRNA fraction; preferably the latter. mRNA was isolated from cells directly in the wells of the tissue culture dish. The cells were lysed and mRNA was bound by the complementary binding of the poly-adenylate tail to a poly-thymidine-associated solid matrix. The bound mRNA was washed several times and the reagents for the reverse transcription (RT) reaction were added. cDNA synthesis in the RT reaction was initiated at random positions along the message by the binding of a random sequence primer (RS). This RS primer has approximately 6-9 random nucleotides at the 3' end to bind sites in the mRNA to prime cDNA synthesis, and a 5' tail sequence of known composition to act as an anchor for PCR amplification in the next step. There is therefore no specificity for the trapped message in the RT step. Alternatively, a poly-dT primer appended with the specific sequences for the PCR may be used. Synthesis of the first strand of the cDNA initiates at the end of each trapped gene. At this point in the procedure, the bound mRNA may be stored (at between about -70° C and about 4° C) and reused multiple times. Such storage is a valuable feature where one subsequently desires to analyze individual clones in more detail. The bound mRNA may also be used to clone the entire transcript using PCR-based protocols.

Specificity for the trapped, fusion transcript is introduced in the next step, PCR amplification. The primers for this reaction are complementary to the anchor sequence of the RS primer and to the selectable marker. Double stranded fragments between a fixed point in the selectable marker gene and various points downstream in the appended transcript sequence are amplified. It is these fragments which will become the substrates for the sequencing reaction. The various end-points along the transcript sequence were determined by the binding of the random primer during the RT reaction. These PCR products were diluted into the sequencing reaction mix, denatured and sequenced using a primer specific for the splice donor sequences of the gene

trap exon. Although, standard radioactively labeled nucleotides may be used in the sequencing reactions, sequences will typically be determined using standard dye terminator sequencing in conjunction with automated 5 sequencers (e.g., ABI sequencers and the like).

Several fragments of various sizes may serve as substrates for the sequencing reactions. This is not a problem since the sequencing reaction proceeds from a fixed point as defined by a specific primer sequence. Typically, 10 approximately 200 nucleotides of sequence were obtained for each trapped transcript. For the PCR fragments that are shorter than this, the sequencing reaction simply 'falls off' the end. Sequences further 3' were then covered by the longer fragments amplified during PCR. One problem is 15 presented by the anchor sequences 'S' derived from the RS primer. When these are encountered during the sequencing of smaller fragments, they register as anomalous dye signals on the sequencing gels. To circumvent this potential problem, a restriction enzyme recognition site is included in the S 20 sequence. Digestion of the double stranded PCR products with this enzyme prior to sequencing eliminates the heterologous S sequences.

5.2.3. Identifying the Tagged Genes by Chromosomal Location

25 Any individually tagged gene may also be identified by PCR using chromosomal DNA as the template. To find an individual clone of interest in the Library arrayed as described above, genomic DNA is isolated from the pooled 30 clones of ES cells as presented in Figure 3. One primer for the PCR is anchored in the gene trap vector, e.g., a puro exon-specific oligonucleotide. The other primer is located in the genomic DNA of interest. This genomic DNA primer may consist of either (1) DNA sequence that corresponds to the 35 coding region of the gene of interest, or (2) DNA sequence from the locus of the gene of interest. In the first case, the only way that the two primers used may be juxtaposed to

give a positive PCR results (e.g., the correct size double-stranded DNA product) is if the gene trap vector has inserted into the gene of interest. Additionally, degenerate primers may be used, to identify and isolate related genes of interest. In the second case, the only way that the two primers used may be juxtaposed to provide the desired PCR result is if the gene trap vector has inserted into the region of interest that contains the primer for the known marker.

10 For example, if one wishes to obtain ES cell clones from the library that contain mutated genes located in a certain chromosomal position, PCR primers are designed that correspond to the puro gene (the puro-anchored primer) and a primer that corresponds to a marker known to be located in
15 the region of interest. Several different combinations of marker primers and primers that are located in the region of interest may also be used to obtain optimum results. In this manner, the mutated genes are identified by virtue of their location relative to sets of known markers. Genes in a
20 particular chromosomal region of interest could therefore be identified. The marker primers could also be designed correspond to sequences of known genes in order to screen for mutations in particular genes by PCR on genomic DNA templates. While this method is likely to be less
25 informative than the RT-PCR strategy described below, this technique would be useful as a alternative strategy to identify mutations in known genes. In addition, primers that correspond to sequence of known genes could be used in PCR reactions with marker-specific primers in order to identify
30 ES cell clones that contain mutations in genes proximal to the known genes. The sensitivity of detection is adequate to find such events when positive clones are subsequently identified as described below in the RT-PCR strategy.

35 5.3. A Sequence Database Identifies Genes Mutated in the Library.

Using the procedures described above, approximately 200

to about 600 bases of sequence from the cellular exons appended to the selectable marker exon (e.g., puro exon in VICTR vectors) may be identified. These sequences provide a means to identify and catalogue the genes mutated in each
5 clone of the Library. Such a database provides both an index for the presently disclosed libraries, and a resource for discovering novel genes. Alternatively, various comparisons can be made between the Library database sequences and any other sequence database as would be familiar to those
10 practiced in the art.

The novel utility of the Library lies in the ability for a person to search the Library database for a gene of interest based upon some knowledge of the nucleic acid or amino acid sequence. Once a sequence is identified, the
15 specific clone in the Library can be accessed and used to study gene function. This is accomplished by studying the effects of the mutation both *in vitro* and *in vivo*. For example, cell culture systems and animal models (i.e., transgenic animals) may be directly generated from the cells
20 found in the Library as will be familiar to those practiced in the art.

Additionally, the sequence information may be used to generate a highly specific probe for isolating both genomic clones from existing data bases, as well as a full length
25 cDNA. Additionally, the probe may be used to isolate the homologous gene from sufficiently related species, including humans. Once isolated, the gene may be over expressed, or used to generate a targeted knock-out vector that may be used to generate cells and animals that are homozygous for the
30 mutation of interest. Such animals and cells are deemed to be particularly useful as disease models (i.e., cancer, genetic abnormalities, AIDS, etc.), for developmental study, to assay for toxin susceptibility or the efficacy of therapeutic agents, and as hosts for gene delivery and
35 therapy experiments (e.g., experiments designed to correct a specific genetic defect *in vivo*).

5.4. Accessing Clones in the Library by a Pooling and Screening Procedure.

An alternative method of accessing individual clones is by searching the Library database for sequences in order to isolate a clone of interest from pools of library clones. The Library may be arrayed either as single clones, each with different insertions, or as sets of pooled clones. That is, as many clones as will represent insertions into essentially every gene in the genome are grown in sets of a defined number. For example, 100,000 clones can be arrayed in 2,000 sets of 50 clones. This can be accomplished by titrating the number of VICTR retroviral particles added to each well of 96-well tissue culture plates. Two thousand clones will fit on approximately 20 such plates. The number of clones may be dictated by the estimated number of genes in the genome of the cells being used. For example, there are approximately 100,000 genes in the genome of mouse ES cells. Therefore, a Library of mutations in essentially every gene in the mouse genome may be arrayed onto 20 96-well plates.

To find an individual clone of interest in the Library arrayed in this manner, reverse transcription-polymerase chain reactions (RT-PCR) are performed on mRNA isolated from pooled clones as presented in Figure 4. One primer for RT-PCR is anchored in the gene trap vector, i.e. a puro exon-specific oligonucleotide. The other primer is located in the cDNA sequence of a gene of interest. The only way that these two sequences can be juxtaposed to give a positive RT-PCR result (i.e. double stranded DNA fragment visible by agarose gel electrophoresis, as will be familiar to anyone practiced in the art) is by being present in a transcript from a gene trap event occurring in the gene of interest.

For example, if one wishes to obtain an ES cell clone with a mutation in the p53 gene, PCR primers are designed that correspond to the puro and p53 genes. If a VICTR trapping vector integrates into the p53 locus and results in the formation of a fusion mRNA, this mRNA may be detected by RT-PCR using these specifically designed primer pairs. The

sensitivity of detection is adequate to find such an event when positive cells are mixed with a large background of negative cells. The individual positive clones are subsequently identified by first locating the pool of 5 clones in which it resides. This process is described in Figure 5. The positive pool, once identified, is subsequently plated at limiting dilution (approximately 0.3 cells/well) such that individual clones may be isolated. To find the one positive event in 50 clones represented by this 10 pool, individual clones are isolated and arrayed on a 96-well plate. By pooling in columns and rows, the positive well containing the positive clone can be identified with relatively few RT-PCR reactions.

In addition to RT-PCR, the pools may be screened by 15 hybridization techniques (see generally Sambrook et al., 1989, Molecular Cloning: H Laboratory Manual 2nd edition, Cold Spring Harbor Press, Cold Spring Harbor, and Current Protocols in Molecular Biology, 1995, Ausubel et al. eds., John Wiley and Sons). Specific PCR fragments are generated 20 from the mutated genes essentially as described above for the sequencing protocols of the individual clones (first-strand synthesis using RT primed by a random or oligo dT primer that is appended to a specific primer binding site). The gene trap DNA is amplified from the primer sets in the puro gene 25 and the specific sequences appended to the RT primer. If this were done with pools, the resulting pooled set of amplified DNA fragments could be arrayed on membranes and probed by radioactive, or chemically or enzymatically labeled, hybridization probes specific for a gene of 30 interest. A positive radioactive result indicates that the gene of interest has been mutated in one of the clones of the positively-labeled pool. The individual positive clone is subsequently identified by PCR or hybridization essentially as outlined above.

35 Alternatively, a similar strategy may be used to identify the clone of interest from multiple plates, or any scheme where a two or three dimensional array (e.g., columns

and rows) of individual clones are pooled by row or by column. For example, 96 well plates of individual clones may be arranged adjacent to each other to provide a larger (or virtual/figurative) two dimensional grid (e.g., four plates may be arranged to provide a net 16x24 grid), and the various rows and columns of the larger grid may be pooled to achieve substantially the same result.

Similarly, plates may simply be stacked, literally or figuratively, or arranged into a larger grid and stacked to provide three dimensional arrays of individual clones. Representative pools from all three planes of the three dimensional grid may then be analyzed, and the three positive pools/planes may be aligned to identify the desired clone. For example, ten 96 well plates may be screened by pooling the respective rows and columns from each plate (a total of 20 pools) as well as pooling all of the clones on each specific plate (10 additional pools). Using this method, one may effectively screen 960 clones by performing PCR on only 30 pooled samples.

The example provided below is merely illustrative of the subject invention. Given the level of skill in the art, one may be expected to modify any of the above or following disclosure to produce insubstantial differences from the specifically described features of the present invention. As such, the following example is provided solely by way of illustration and is not included for the purpose of limiting the invention in any way whatsoever.

6.0. EXAMPLES

6.1. Use of VICTR Series Vectors to Construct a Mouse ES cell Gene Trap Library

VICTR 3 was used to gather a set of gene trap clones. A plasmid containing the VICTR 3 cassette was constructed by conventional cloning techniques and designed to employ the features described above. Namely, the cassette contained a PGK promoter directing transcription of an exon that encodes the puro marker and ends in a canonical splice donor

sequence. At the end of the puromycin exon, sequences were added as described that allow for the annealing of two nested PCR and sequencing primers. The vector backbone was based on pBluescript KS+ from Stratagene Corporation.

- 5 The plasmid construct linearized by digestion with Sca I which cuts at a unique site in the plasmid backbone. The plasmid was then transfected into the mouse ES cell line AB2.2 by electroporation using a BioRad Genepulser apparatus. After the cells were allowed to recover, gene trap clones
10 were selected by adding puromycin to the medium at a final concentration of 3 μ g/mL. Positive clones were allowed to grow under selection for approximately 10 days before being removed and cultured separately for storage and to determine the sequence of the disrupted gene.
- 15 Total RNA was isolated from an aliquot of cells from each of 18 gene trap clones chosen for study. Five micrograms of this RNA was used in a first strand cDNA synthesis reaction using the "RS" primer. This primer has unique sequences (for subsequent PCR) on its 5' end and nine
20 random nucleotides or nine T (thymidine) residues on its 3' end. Reaction products from the first strand synthesis were added directly to a PCR with outer primers specific for the engineered sequences of puromycin and the "RS" primer. After amplification, an aliquot of reaction products were subject
25 to a second round of amplification using primers internal, or nested, relative to the first set of PCR primers. This second amplification provided more reaction product for sequencing and also provided increased specificity for the specifically gene trapped DNA.
- 30 The products of the nested PCR were visualized by agarose gel electrophoresis, and seventeen of the eighteen clones provided at least one band that was visible on the gel with ethidium bromide staining. Most gave only a single band which is an advantage in that a single band is generally
35 easier to sequence. The PCR products were sequenced directly after excess PCR primers and nucleotides were removed by filtration in a spin column (Centricon-100, Amicon). DNA was

added directly to dye terminator sequencing reactions (purchased from ABI) using the standard M13 forward primer a region for which was built into the end of the puro exon in all of the PCR fragments. Thirteen of the seventeen clones
5 that gave a band after the PCR provided readable sequence. The minimum number of readable nucleotides was 207 and some of the clones provided over 500 nucleotides of useful sequence.

Sample data from this set of clones is presented in
10 Figure 6. Only a portion of sequence (nucleotide or putative amino acid) for 9 Library clones obtained by the methods described in this invention are presented. Under each sequence fragment in the figure is aligned a homologous sequence that was identified using the BLAST (basic local
15 alignment search tool) search algorithm (Altschul et al., 1990, J. Mol. Biol. 215:403-410).

In addition to known sequences, many new genes were also identified. Each of these sequences is labeled "OST" for "Omnibank Sequence Tags." OMNIBANK™ shall be the trademark
20 name for the Libraries generated using the disclosed technology.

These data demonstrate that the VICTR series vectors may efficiently trap genes, and that the procedures used to obtain sequence are reliable. With simple optimization of
25 each step, it is presently possible to mutate every gene in a given population of cells, and obtain sequence from each of these mutated genes. The sample data provided in this example represents a small fraction of an entire Library. By simply performing the same procedures on a larger scale (with
30 automation) a Library may be constructed that collectively comprises and indexes mutations in essentially every gene in the genome of the target cell.

Additional studies have used both VICTR 3 and VICTR 20. Like VICTR 3, VICTR 20 is exemplary of a family of vectors
35 that incorporate two main functional units: a sequence acquisition component having a strong promoter element (phosphoglycerate kinase 1) active in ES cells that is fused

to the puromycin resistance gene coding sequence which lacks a polyadenylation sequence but is followed by a synthetic consensus splice donor sequence (PGKpuroSD); and 2) a mutagenic component that incorporates a splice acceptor sequence fused to a selectable, colorimetric marker gene and followed by a polyadenylation sequence (for example, SA β geopA or SAIRES β geopA). Also like VICTR 3, stop codons have been engineered into all three reading frames in the region between the 3' end of the selectable marker and the splice donor site. A diagrammatic description of structure and functions of VICTRs 3 and 20 is provided in Figure 7.

When VICTRs 3 and 20 were used in the commercial scale application of the presently disclosed invention, over 3,000 mutagenized ES cell clones were rapidly engineered and obtained. Sequence analysis obtained from these clones has identified a wide variety of both previously identified and novel sequences. A representative sampling of previously known genes that were identified using the presently described methods is provided in Figure 8. The power of the presently described invention as a genomics resource becomes apparent when one considers that the genes listed in Figure 8 were obtained and identified in less than a year whereas the references associated with the identification of the known genes span a period of roughly two decades. More importantly, the majority of the sequences thus far identified are novel, and, because of the functional aspects of the presently described ES cell system, the cellular and developmental functions of these novel sequences can be rapidly established.

30

7.0. Reference to Microorganism Deposits

The following plasmids have been deposited at the American Type Culture Collection (ATCC), Rockville, MD, USA, under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure and Regulations thereunder (Budapest Treaty) and are thus maintained and made available according

to the terms of the Budapest Treaty. Availability of such plasmids is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The deposited cultures have been assigned the indicated ATCC deposit numbers:

	<u>Plasmid</u>	<u>ATCC No.</u>
	plex	97748
	pExonII	97749
10	ppuro7	97750
	ppuro5	97751
	ppuro11	97752
	ppuro10	97753

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

30

35

MICROORGANISMS	
Optional Sheet in connection with the microorganism referred to on page <u>40</u> , lines <u>5-25</u> of the description *	
A. IDENTIFICATION OF DEPOSIT * Further deposits are identified on an additional sheet *	
Name of depositary institution * American Type Culture Collection	
Address of depositary institution (including postal code and country) * 12301 Parklawn Drive Rockville, MD 20852 US	
Date of deposit * <u>October 9, 1996</u> Accession Number * <u>97748</u>	
B. ADDITIONAL INDICATIONS * (leave blank if not applicable). This information is continued on a separate attached sheet	
C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE * (if the indications are for all designating States)	
D. SEPARATE FURNISHING OF INDICATIONS * (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later * (Specify the general nature of the indications e.g., "Accession Number of Deposit")	
E. <input type="checkbox"/> This sheet was received with the International application when filed (to be checked by the receiving Office) <div style="text-align: right;">_____ (Authorized Officer)</div> <input type="checkbox"/> The date of receipt (from the applicant) by the International Bureau " was _____ (Authorized Officer)	

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

12301 Parklawn Drive
Rockville, MD 20852
US

Accession No.

Date of Deposit

97749	October 9, 1996
97750	October 9, 1996
97751	October 9, 1996
97752	October 9, 1996
97753	October 9, 1996

CLAIMS

What is claimed is:

1. A library of cultured eucaryotic cells made by a process comprising the steps of:
 - 5 a) treating a first group of cells to stably integrate a first vector that mediates the splicing of a foreign exon internal to a cellular transcript;
 - b) treating a second group of cells to stably integrate a second vector that mediates the splicing of a foreign exon
10 5' to an exon of a cellular transcript; and
 - c) selecting for transduced cells that express the products encoded by the foreign exons.
2. A library according to claim 1 wherein said treating
15 is transfection.
3. A library according to claim 1 wherein said treating is by infection.
- 20 4. A library according to claim 1 wherein said treating is by retrotransposition.
5. A library according to any one of claims 1 through 4 wherein said cells are animal cells.
- 25 6. A library according to claim 5 wherein said animal is mammalian.
7. A library according to claim 6 wherein said cells
30 are rodent cells.
8. The use of a mutated cell from a library according to claim 6 to generate a non-human transgenic animal.
- 35 9. A vector for replacing the 3' end of an animal cell transcript with a foreign exon, comprising:
 - a) a selectable marker;

- b) a splice acceptor site operatively positioned 5' to the initiation codon of said selectable marker;
- c) a polyadenylation site operatively positioned 3' to said selectable marker;
- 5 d) said vector not comprising a promoter element operatively positioned 5' of the coding region of said selectable marker; and
- e) said vector not comprising a splice donor sequence operatively positioned between the 3' end of the
- 10 coding region of said selectable marker and said polyadenylation site.

10. A vector for inserting foreign mutagenic polynucleotide sequence internal to animal cell transcripts, comprising:

- a) a foreign exon;
- b) a splice acceptor sequence operatively positioned 5' to the foreign exon;
- c) a splice donor site operatively positioned 3' to said foreign exon;
- 20 d) a sequence comprising a nested set of stop codons in each of the three reading frames located between the 3' end of said foreign exon and said splice donor site;
- 25 e) said vector not comprising a polyadenylation site operatively positioned 3' to said foreign exon; and
- f) said vector not comprising a promoter element operatively positioned 5' to the coding region of said foreign exon.

30

11. A vector for attaching a foreign exon upstream from the 3' end of an animal cell transcript, comprising:

- a) a selectable marker;
- b) a promoter element operatively positioned 5' to said selectable marker;
- 35 c) a splice donor site operatively positioned 3' to said selectable marker; and

- d) said vector not comprising a transcription terminator or polyadenylation site operatively positioned relative to the coding region of said selectable marker; and
- 5 e) said vector not comprising a splice acceptor site operatively positioned between said promoter element and the initiation codon of said selectable marker.
- 10 12. A vector according to claim 11 wherein said vector additionally comprises a foreign mutagenic polynucleotide sequence located upstream from said promoter.
- 15 13. A vector according to claim 12 wherein said vector additionally comprises a splice acceptor operatively positioned upstream from said foreign mutagenic polynucleotide sequence.
- 20 14. A vector according to claim 13 wherein said foreign mutagenic polynucleotide sequence comprises a polyadenylation site.
- 25 15. A vector according to claim 14, wherein said foreign mutagenic polynucleotide sequence additionally comprises stop codons in all three reading frames.
- 30 16. A vector according to claim 12 in which a first recombinase recognition sequence is present upstream from said promoter and a second recombinase recognition sequence is present downstream from said promoter.
17. A vector according to any one of claims 9, 10, or 11 wherein said vector is a viral vector.
- 35 18. A vector according to claim 17 wherein said viral vector is a retroviral vector.

19. The use of a vector according to claim 9 to produce a library of mutated animal cells.

20. The use of a vector according to claim 10 to produce mutated animal cells.

21. The use of a vector according to claim 11 to produce mutated animal cells.

22. The use of a vector according to claim 11 to effect homologous recombination in an animal cell.

23. A stably transduced animal cell that incorporates a vector according to claim 16.

15

24. A method of deleting a region of vector DNA from a cell according to claim 23, comprising:

- a) providing a recombinase activity to the cell; and
- b) selecting for cells that lack the desired region of vector DNA.

25. A method of adding a region of DNA to a cell according to claim 23, comprising:

- a) introducing the DNA to be added into the cell;
- a) providing a recombinase activity to the cell; and
- b) selecting for cells that incorporate the added DNA.

26. A method of effecting the inducible expression of a desired gene, comprising:

- a) providing a cell according to claim 23 with a recombinase gene that is expressed by an inducible promoter; and
- b) inducing said inducible promoter.

27. A method of gene discovery comprising:

- a) adding a foreign polynucleotide to a population of target cells such that the foreign

polynucleotide is inserted throughout the genomes of the target cells; and

b) activating control elements encoded by the foreign polynucleotides that activate or repress the expression of target cell genes that flank the integrated foreign polynucleotides, and identifying the regions of the target cell genome into which the foreign polynucleotides have integrated.

10 28. A library of cultured animal cells that stably integrate vectors according to claims 10 or 11.

15

20

25

30

35

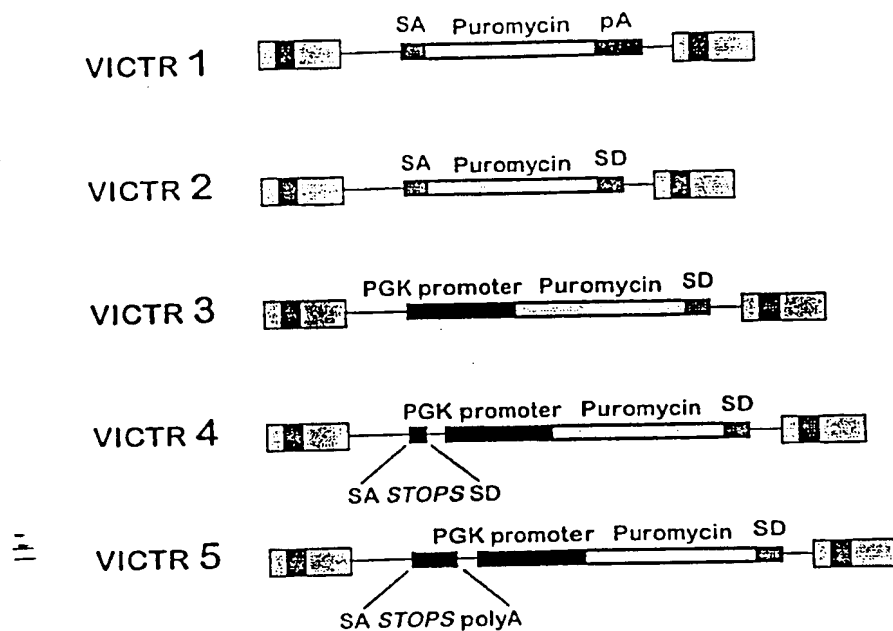


Figure 1

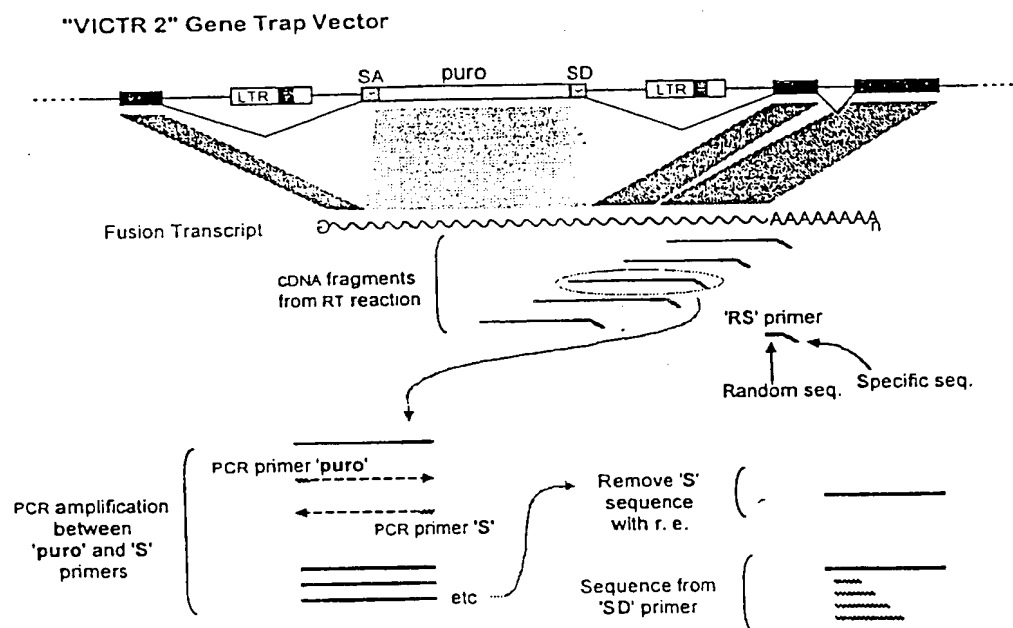


Figure 2

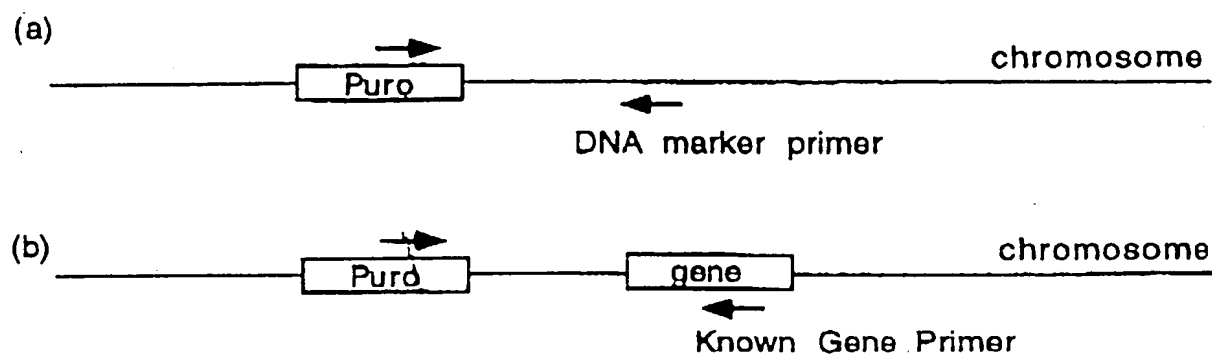


Figure 3

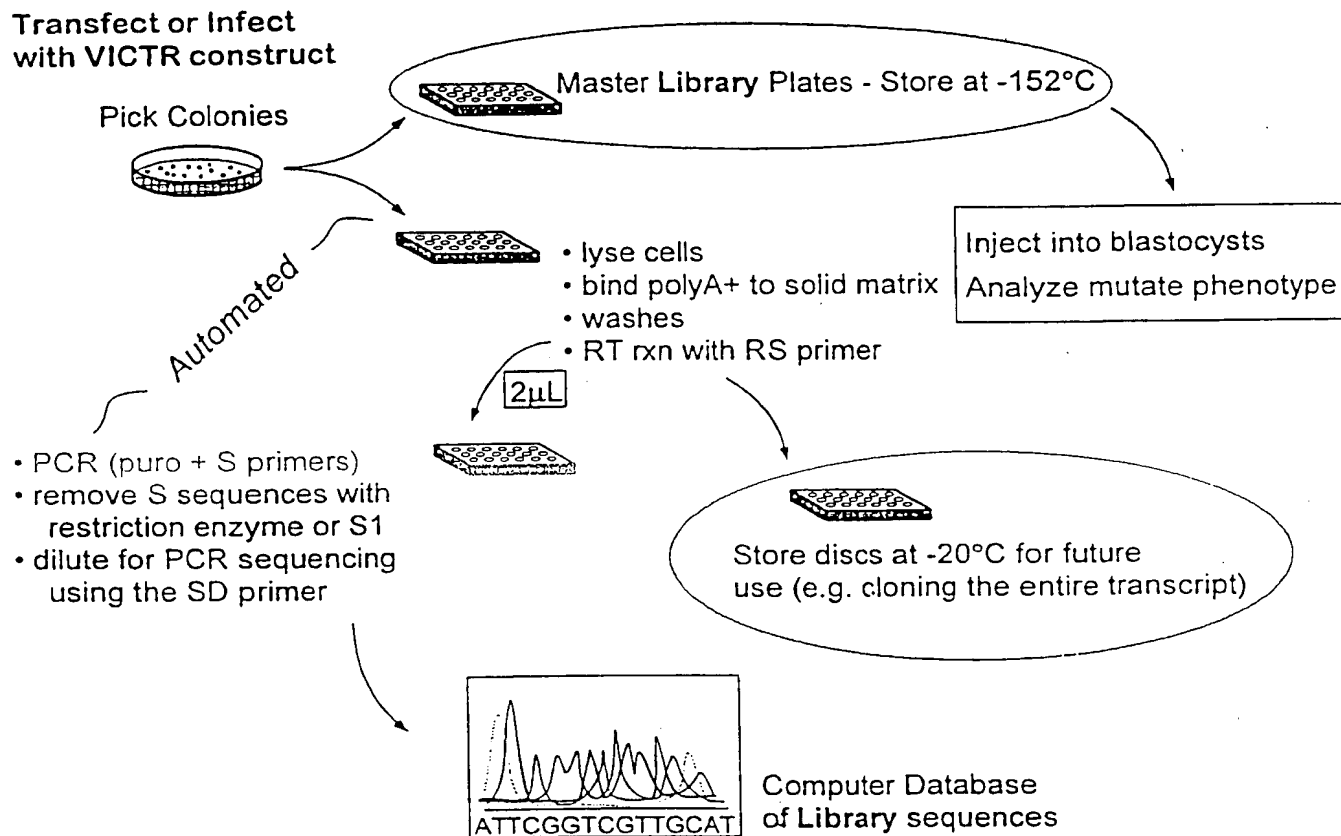
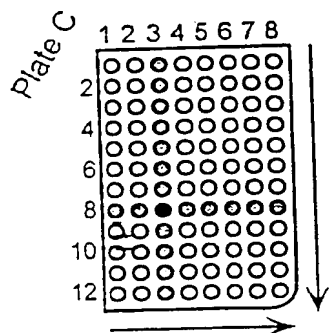


Figure 4

Identify Positive Pool

To screen all mouse genes (~100,000) with 5-fold redundancy would require about 50 plates of 96-wells (at 100 clones/well).



① Pool columns

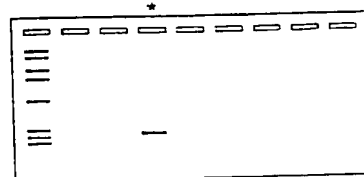
(100 clones/well)(12 wells) = 1200 clones per rxn.
(50 plates)(8 columns/plate) = 400 initial rxns.

Plate C, column 3 was positive.

② Pool positive rows only

8 rxns per positive plate

Row 8 was positive.

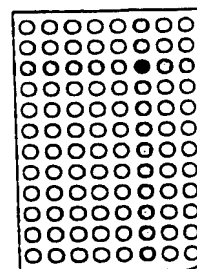


Identify Positive Clone

The pool on plate C, column 3, row 8 is thawed and plated as single clones:



① Pick 96 colonies into two 96-well plates



② Pool columns (16) and rows (24) for 40 rxns to identify positive clone.

Figure 5

OST1:	248	TTTATATAATATTTAATTGTTTTACTGGGGTATATATGTGTGAAGAGGACTTCT	302
rat GABA rho3:	1547	TTTACATAATATTTAATTGTTTTACTGGGGTATATATGTGTGAAGAGGACTTTT	1601
OST2:	56	ACCGTTGCGGAGGCTCACGTTTCTCAGATAGTACATCAGGTGTCATCGNTGTCAGAAGGT	115
mouse TCR-ATF1:	75	ACCGTTGCGGGGCTCACGTTTCTCAGATAGTACATCAGGTGTCATCGTTATCAGAAAGT	134
OST3:	58	GIGMHAGLHERDRKTVEELFXNCKVQVLIATSTLAWGVNFPahlVIiKGTEYYDGKTRR	237
Yeast ORF G9365:	1430	GIGLHHAGLVQKDRSISHQLFQKNKIQILIATSTLAWGVNLPahlVIiKGTQFFDAKTEG	1489
OST4:	137	GCGCAGAAGTGGTNTCTGGAANTTTNTCCGCCNCCATCCAGTCTATTAAATTGTTGACNGGA	196
seq. from US patent 5470724:	166	GCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAAATTGTTGCCGGGA	225
OST5:	108	TCWIRLGT*RXVGASLEYEYIRAS	179
mouse wnt-5A protein precursor:	250	TCW++L R VG +L+ +Y A+	273
OST6:	78	CITATATGGCTACGGCGGCTTCAACATCTCCATTACACCCAACTACAGCGTGTCCAGGCT	137
human prolyl endopeptidase:	1407	CITATATGGCTATGGCGGCTTCAACATATCCATCACACCCAACTACAGTGTTCAGGCT	1466
OST7:	109	AAAGCATGTAGCAGTTGTAGGACACACTAGACGAGAGCACCAGATCTCATGTGGGTGGT	168
mouse 45S pre rRNA:	1604	AAAGCATGTAGCAGTTGTAGGACACACTAGACGAGAGCACCAGATCTCATGTGGGTGGT	1663
OST8:	161	TGGATGCAGNCTACCACTGTGTGGCTGCCCTATTTTACCTCAGTGCCCTCAGTTCTGGAAG	220
rat MAL:	306	TGGATGCAGCCTACCACTGTGTGGCTGCCCTSTTTTACCTCAGTGCCCTCAGTCTGGAAG	365
OST9:	103	ACCTGATGTATTATCCGTGGCCTGCAGAAGTCCAGAAAATACAGACCAAAGTCAACCAGTA	162
mouse malic enzyme:	1666	ACCTGATGTATTATCCGTGGCCTGCAGAAGTCCAGAAAATACAGACCAAAGTCAACCAGTA	1725

Figure 6

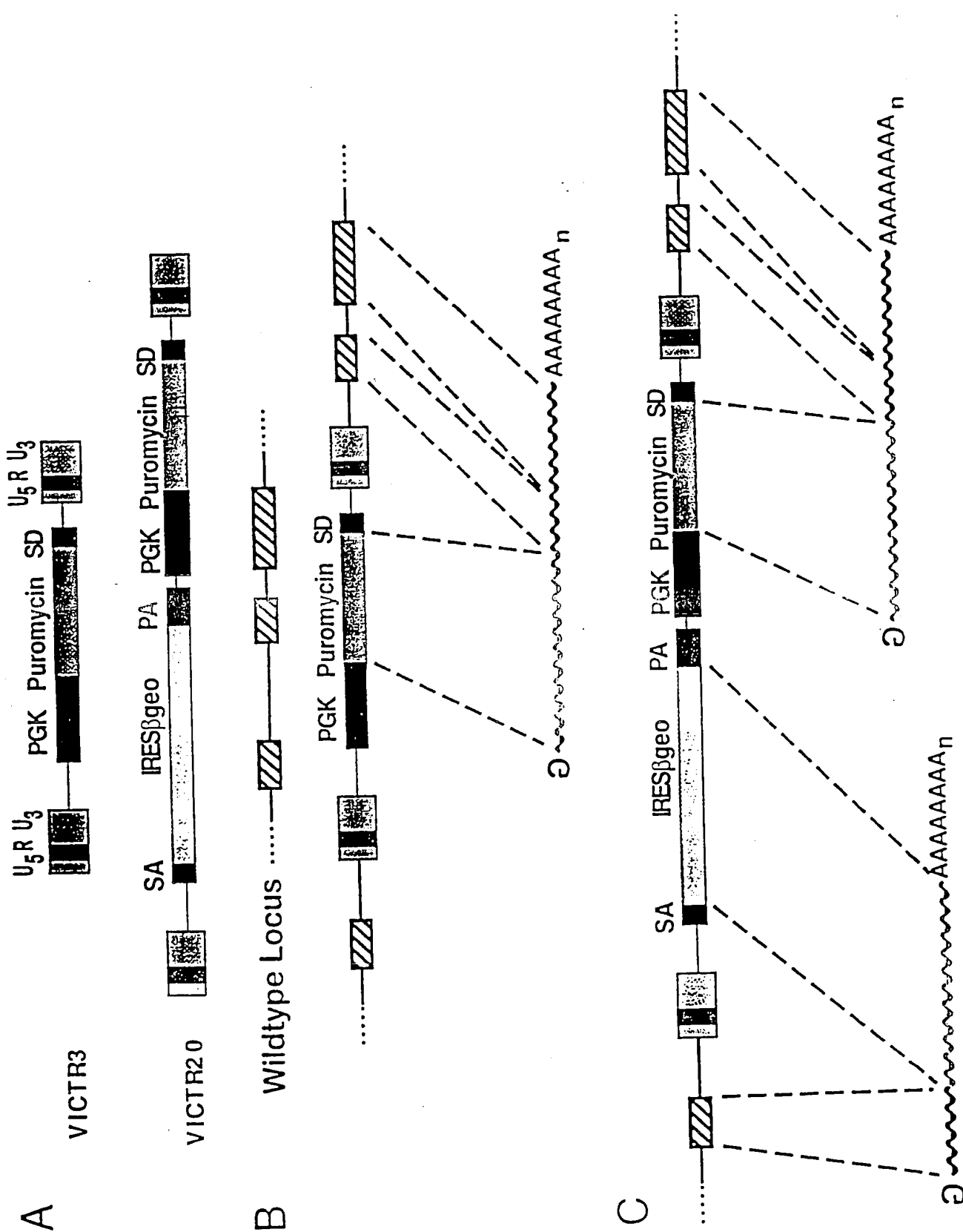


Figure 7

Figure 8

The following table includes 516 OSTs. OSTs with hit into protein and Genbank patented sequences have been removed as well as sequence with repetitive elements hits.			
Genbank	Accession	Size	Sequence Description
OST4	U01007445	5,000-113	Mus musculus mG102.11 Soares mouse p1507.5.113 Mus musculus cDNA clone
OST5	U01007446	2,000-11	Mus musculus Mouse mRNA for retinal cyclic-GMP phosphodiesterase
OST6	U01007447	5,900-148	gamma-subunit (GMP-PDE) (EC 3.1.4.17)
OST7	U01007448	1,000-42	Mus musculus Mouse mRNA
OST8	U01007449	1,900-173	Complete cds m150b06.r1 Soares mouse
OST9	U01007450	1,900-173	Mus musculus m150b06.r1 Soares mouse
OST10	U01007451	7,500-71	Mus musculus m150b06.r1 Soares mouse
OST11	U01007452	1,000-106	Mus musculus Mouse mRNA for squalene synthase
OST12	U01007453	1,000-106	Mus musculus M. musculus T cell receptor alpha chain variable region (V-alpha)
OST13	U01007454	1,800-70	Mus musculus mouse alpha-amylose-2
OST14	U01007455	4,000-34	Gene: pancreatic m1na
OST15	U01007456	1,400-145	Rattus norvegicus Rat cytochrome P450 II A3 (CYP2A3) gene, complete cds
OST16	U01007457	1,400-145	Mus musculus m917d10.r1 Soares mouse
OST17	U01007458	1,500-45	embryo NMEL1.5 14.5 Mus musculus cDNA clone 42631.5
OST18	U01007459	2,600-37	Mus musculus House mouse; Musculus domesticus Postnatal (0 day) Brain mRNA for Ca2+ dependent activator protein for secretion, complete cds
OST19	U01007460	7,500-112	Mus musculus m152c11.r1 Soares mouse
OST20	U01007461	1,000-126	lymph node NM14H Mus musculus cDNA clone 643028.5 similar to TR:G294850
OST21	U01007462	1,700-31	G294850 ALPHA-MUSCLE ACTIN
OST22	U01007463	1,700-31	Rattus norvegicus Rat TH-4 gene for fibroblast tropomyosin 4
OST23	U01007464	1,700-31	Mus musculus R. musculus 191-V42(70/73)
OST24	U01007465	1,800-178	Mus musculus m146105.r1 Soares mouse
OST25	U01007466	1,800-178	lymph node NM14H Mus musculus cDNA clone 642465.5
OST26	U01007467	7,300-40	Mus musculus m056003.r1 Life Tech mouse embryo 8.5dpc 10664019 Mus musculus cDNA clone 557573.5 similar to SM:YA36_SC100 009713 HYPOTHETICAL 37.7 KB PROTEIN C18011.06 IN CH11
OST27	U01007468	4,000-111	Rattus sp. EST110153 Rattus sp. cDNA 5' end
OST28	U01007469	8,000-154	Homo sapiens 2151b07.s1 Soares pregnant uterus NM14H Homo sapiens cDNA clone 505429.3 similar to TR:G632498 G632498 CLEAVAGE STIMULATION FACTOR 7KDa SUBUNIT
OST29	U01007470	2,000-145	Homo sapiens Human mRNA for KIAA0242 gene, partial cds
OST30	U01007471	3,100-161	Homo sapiens Human sc12 mRNA for RHA binding protein SC12, complete cds
OST31	U01007472	1,200-52	Homo sapiens ang6h09.r1 Stratyene HeLa cell s1 971216 Homo sapiens cDNA clone 563201.5
OST32	U01007473	4,000-148	Mus musculus m019c06.r1 Life Tech mouse embryo 10.5dpc 1065016 Mus musculus cDNA clone 556906.5 similar to G0105277 Mouse hexokinase mRNA, complete cds (MOUSE)
OST33	U01007474	8,100-143	Rattus norvegicus Rat mRNA for ribophorin I
OST34	U01007475	4,800-107	Homo sapiens similar to glutamyl-tRNA synthetase
OST35	U01007476	4,800-38	Rattus sp. EST106933 Rattus sp. cDNA 5' end similar to Synapsin I
OST36	U01007477	1,800-81	Mus musculus m101b09.r1 Soares mouse
OST37	U01007478	1,200-91	embryo NMEL1.5 14.5 Mus musculus cDNA clone 441205
OST38	U01007479	1,500-141	Mus musculus Mouse 4.5S mRNA gene
OST39	U01007480	1,500-141	Mus musculus m17411.r1 Soares mouse
OST40	U01007481	1,500-141	embryo NMEL1.5 14.5 Mus musculus cDNA clone 418164.5

05T562	9b x61433	7.6e-68	97%	HA01-UH01 INTERGENIC REGION. [1] Mus musculus H. musculus mRNA for sodium/potassium ATPase beta subunit
05T568	9b AA007930	1.5e-31	67%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T573	9b AA111278	2.1e-147	92%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T574	9b AA130347	1.2e-103	85%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T575	9b 142055	4.0e-60	75%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T577	9b AA020429	2.1e-91	92%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T581	9b W06552	2.0e-90	80%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T582	9b D17695	1.9e-218	91%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T591	9b 143326	3.6e-103	85%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T593	9b W07077	3.4e-117	98%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T594	9b x94616	2.6e-142	87%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T595	9b W07137	7.0e-51	86%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T598	9b x53476	2.2e-235	98%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T600	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T601	9b W55702	1.2e-71	85%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T602	9b AA117282	1.0e-78	83%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T603	9b AA007930	2.1e-60	91%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T604	9b W11817	1.5e-95	86%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T605	9b AA117282	1.0e-78	83%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T606	9b AA007930	2.1e-60	91%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T607	9b AA007930	2.1e-60	91%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T608	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T609	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T610	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T611	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T612	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T613	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T614	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T615	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T616	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T617	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T618	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T619	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T620	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T621	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T622	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T623	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T624	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T625	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T626	9b W070494	1.0e-188	96%	embryo NMEL13.5 14.5 Mus musculus cDNA clone 477500 5' similar to gh-202809
05T627	9b W070494	1.0e-188		

05R1096	gb D87077	7.7e-112	881	ICX31.11 minA, complete cds Homo sapiens human minA for KIAA0210 gene, partial cDNA clone 321647.3
05R1105	gb W44423	1.0e-66	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1116	gb X22249	3.1e-36	801	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1117	gb X22485	1.2e-88	941	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1118	gb X22486	1.2e-128	911	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1152	gb W79707	6.4e-109	101	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1155	gb U01024	2.1e-65	951	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1165	gb U54649	1.9e-184	971	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1179	gb AA008986	8.5e-84	911	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1186	gb U09840	2.8e-70	961	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1192	gb U08470	5.3e-127	961	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1207	gb AA063763	1.3e-56	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1223	gb AA002931	1.5e-189	991	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1226	gb U37353	7.5e-279	951	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1233	gb W08703	6.3e-91	981	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1234	gb U085904	1.1e-180	981	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1241	gb U09850	4.8e-184	921	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1247	gb AA051266	4.7e-126	971	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1265	gb X36373	6.8e-103	961	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1267	gb U081271	1.6e-32	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1269	gb U019977	4.0e-130	811	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1274	gb W14634	2.1e-139	851	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1294	gb J03583	6.0e-69	951	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1339	gb W97190	3.1e-118	971	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1341	gb W09611	7.0e-142	931	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1354	gb W14401	2.2e-64	851	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1359	gb W06262	5.1e-35	921	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1369	gb W062550	1.1e-109	971	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1376	gb X63615	1.1e-33	921	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1389	gb X15521	6.5e-57	931	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1410	gb W59607	4.8e-77	891	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1419	gb W10923	1.2e-122	971	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1432	gb W09809	1.7e-145	941	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1453	gb C18312	5.6e-92	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1457	gb W78851	5.9e-45	801	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1470	gb J05504	1.5e-136	951	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1478	gb U57821	0.4e-55	801	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1508	gb J024681	4.3e-235	911	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1592	gb U010701	3.6e-109	981	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1593	gb AA007483	3.5e-197	981	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1599	gb T51184	1.3e-60	911	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1604	gb J34260	3.7e-196	911	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1608	gb J03306	1.8e-164	941	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1620	gb W18420	4.7e-37	881	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1623	gb U23769	7.2e-195	921	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1654	gb W06270	2.3e-168	961	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1656	gb W117514	9.7e-150	891	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1658	gb J03785	2.2e-67	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1667	gb W14636	1.7e-39	941	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1681	gb W10611	1.3e-89	851	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1683	gb W09922	4.0e-109	861	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1684	gb U57139	7.9e-78	931	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1689	gb AA010090	7.9e-158	961	Homo sapiens minA, complete cds Homo sapiens minA, complete cds
05R1693	gb AA010795	5.1e-143	971	

[illegible]

Figure 8 cont'd.

OST2829	gb AA02649	7.7e-90	941	LPS-binding protein embryo NMH13.5 14.5 Mus musculus cDNA clone 426106 5' similar to SW:HS5_HUMAN 000587 SENON PROTEIN M555. 11
OST2814	gb U07632	1.4e-222	971	Mus musculus Mus musculus H-terminal asparagine amidohydrolase (Hs011)
OST2815	gb AA060795	2.1e-89	971	Mus musculus m37405.01 1 Soares mouse cDNA clone 1355 5' similar to p22115.5 Mus musculus cDNA clone p22115.5 similar to WP:F42110.4
OST2819	gb AA163971	6.0e-61	701	Mus musculus m31001.11 Life Tech mouse embryo 13 Sdpc 1066011 Mus musculus cDNA clone 613992 5'
OST2842	gb U01515	6.1e-64	911	Mus musculus m09410.11 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 367866 5' similar to gb:U07151 ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 3
OST2877	gb J001583	1.3e-66	911	Rattus norvegicus Rat clathrin heavy chain mRNA, complete cds
OST2803	gb U04850	4.0e-75	911	Mus musculus m62602.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 353067 5' similar to gb:U11248 Mus musculus C57BL/6J ribosomal protein 528 mRNA, complete (H005)
OST2892	gb U07778	1.4e-125	911	embryo NMH13.5 14.5 Mus musculus cDNA clone 422538 5' similar to embryo NMH13.5 14.5 Mus musculus cDNA clone 422538 5' similar to
OST2897	gb U11047	7.9e-132	971	POLYPEPTIDE VII-LIVER/HEART Mus musculus m47810.01 Soares mouse p1HMF19.5 Mus musculus cDNA clone 316819 5'
OST2909	gb AA166258	8.9e-120	961	Mus musculus m49609.01 Life Tech mouse embryo 13 Sdpc 1066011 Mus musculus cDNA clone 614856 5' cDNA Mus musculus Mus musculus cDNA nucleolar phosphoprotein p102 mRNA, complete cds
OST2911	gb U073478	1.4e-117	861	Mus musculus Mus musculus ANK alpha p190 transcript mRNA, complete cds Mus musculus Mus musculus embryonal carcinoma F9 cell cDNA, 91E10
OST2914	gb U012236	4.0e-136	951	Mus musculus m65506.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 391379 5'
OST2916	gb U07002	1.4e-67	921	Homo sapiens human lymphocyte mRNA for TFIID subunit p22, complete cds
OST2921	gb U05740	8.4e-106	981	Mus musculus m37401.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 407209 5'
OST2922	gb U050544	8.4e-135	881	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 407209 5'
OST2923	gb U085631	3.2e-108	971	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 407209 5'
OST2926	gb U055561	6.3e-164	941	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 373968 5'
OST2929	gb U075735	3.0e-92	921	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 390983 5'
OST2934	gb U082904	1.8e-75	931	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 390983 5'
OST2940	gb AA154335	1.4e-114	971	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 390983 5'
OST2942	gb U034802	1.4e-91	961	Mus musculus m47260.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 390983 5'
OST2918	gb AA108292	5.1e-32	811	Rattus norvegicus R570035 Rat lambda ZAP1 library (C.D. Hamel) Mus norvegicus cDNA clone pC03 5' similar to ADP-ribosylation factor (ARF)-like protein
OST2953	gb U10606	1.8e-97	981	Mus musculus m44101.01 Soares mouse p1HMF19.5 Mus musculus cDNA clone 313557 5' similar to p1H35037R J06737P Arpase inhibitor protein precursor, mitochondrial - rat 11
OST2974	gb AA019172	3.1e-137	971	Mus musculus m31001.11 Soares mouse
OST2963	gb U04744	4.2e-31	801	embryo NMH13.5 14.5 Mus musculus cDNA clone 479149 5' similar to WP:F45E12.4 CE02740
OST2971	gb AA120487	9.2e-107	101	Homo sapiens zaf9c08.01 Soares fetal lung NMH19W Homo sapiens cDNA clone 298766 5'
OST2974	gb U03553	2.6e-102	181	Mus musculus m12107.01 Beddington mouse embryonic region Mus musculus cDNA clone 53773 5' similar to SW:YH5_YEAST J08219 HYPOTHETICAL 44.2 KD PROTEIN IN SC02-MHF1 INTENSE REGION
OST2977	gb U07755	6.1e-164	971	Rattus norvegicus Rattus norvegicus neuroglycan C precursor mRNA, complete cds
OST2980	gb AA206420	1.2e-71	851	Mus musculus M.musculus mRNA MSI Homo sapiens zaf1002.01 Stragene neuroepithelium (9377) Homo sapiens cDNA clone 4583 5' similar to TRIGLYCERIN TAG SEQUENCE TAG
OST2983	gb U02906	1.1e-119	911	Mus musculus m31012.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 355942 5' similar to PIN:S44900 S44900 22652.10 protein - Carnitine acetyltransferase
OST2987	gb AA027603	2.3e-134	961	Mus musculus m12001.01 Soares mouse p1HMF19.5 Mus musculus cDNA clone 463273 5'
OST2988	gb U052129	2.2e-52	731	Mus musculus domesticus Mouse testis-specific mRNA p86.2
OST2989	gb AA152050	1.3e-46	781	Homo sapiens z148b12.01 Soares pregnant uterus NMH19 Homo sapiens cDNA clone 505151 5' similar to gb:M00356_cds1 TRANSCRIPTION FACTOR BTF3 (HUMAN)
OST2991	gb AA003171	8.4e-151	931	Mus musculus m956109.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 437057 5' similar to gb:M24194 GUANINE NUCLEOTIDE-BINDING PROTEIN UETA SUBUNIT-LIKE PROTEIN (HUMAN); BTF3 (HUMAN)
OST2994	gb U051546	1.9e-51	811	Homo sapiens y972h12.01 Homo sapiens cDNA clone 38905 5' similar to SP-VIL1-CHICK P02640
OST2996	gb U09921	1.6e-82	101	Calcium-binding protein A11 Mus musculus Mus musculus cDNA, MUSG01203, clone AC0315
OST2998	gb U019012	3.2e-48	101	Mus musculus Mus musculus cDNA, MUSG01203, clone AC0315
OST2999	gb U07502	1.3e-169	971	Mus musculus Mus musculus lens major intrinsic protein (MIP) mRNA, complete cds
OST3001	gb AA103385	1.9e-162	981	Mus musculus m02102.01 Life Tech mouse embryo 13 Sdpc 1066011 Mus musculus cDNA clone 554427 5' similar to gb:215030.01 MYOSIN REGULATORY LIGHT CHAIN 2, VENTRICULAR (HUMAN); gb:X65979 M.musculus PLKLC-A mRNA (for myosin light chain 2 (HOUSE))
OST3011	gb AA035005	1.2e-98	991	Mus musculus m153a10.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 467226 5' similar to PIN:S28237 S28237 NA01 dehydrogenase
OST3017	gb U050908	4.8e-123	921	Mus musculus m310e02.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 476762 5' similar to SW:AP17-MAT Q00380 CLATHRIN COAT ASSEMBLY PROTEIN AP17
OST3018	gb U00277	2.2e-235	991	Mus musculus House DIA for Small GTP-binding protein 510, exon2 and complete cds
OST3032	gb U07185	2.1e-76	991	Mus musculus Mus musculus CTP synthetase homolog (CTPH) mRNA, complete cds
OST3035	gb U08631	1.8e-115	901	Mus musculus Mus musculus large ribosomal subunit protein mRNA, complete cds
OST3049	gb U090976	4.5e-34	741	Mus musculus m1H405.01 Soares mouse embryo NMH13.5 14.5 Mus musculus cDNA clone 421017 5'

05T138J	gb 279416	1.1e-114	921	Mus musculus H. musculus OUL1 mRNA for outer dense fiber protein of sperm tails
05T138S	gb 083024	1.4e-75	861	homo sapiens similar to T cell-specific PAL
05T139Z	gb w05918	4.7e-139	921	Mus musculus m080109.1 Soares mouse p1NH19.5 Mus musculus cDNA clone 301953.5
05T139A	gb w41666	1.1e-110	931	Mus musculus m082001.1 Soares mouse clone 374881.5 similar to SW:GUS5 MOVIN P30670 GUANINE NUCLEOTIDE-BINDING PROTEIN G11/G15/G101 GAMMA-5 SUBUNIT. [11]
05T1300	gb 062403	2.1e-180	901	Mus musculus Mus musculus ubiquitin conjugating (m1yme labe4) mRNA, complete cds
05T1201	gb 052851	6.0e-54	901	homo sapiens human fetal brain cDNA 5'-end GGU-070103
05T1205	gb w40803	1.9e-173	991	Mus musculus m039007.1 Soares mouse p1NH19.5 Mus musculus cDNA clone 350893.5
05T1308	gb w423458	2.0e-119	901	Mus musculus Mus musculus endogenous retroviruslike D-36 (distantly related to MuLV) LTR
05T13516	gb 114441	5.4e-177	901	Rattus norvegicus Rat phosphatidylethanolamine N-methyltransferase mRNA, complete cds
05T13517	gb AA015044	5.3e-114	971	Mus musculus m02310.1 Soares mouse placenta 410BP11.5 14.5 Mus musculus cDNA clone 443371.5
05T13518	gb AA061165	6.3e-99	911	Mus musculus m311105.1 Soares mouse embryo m0611.5 14.5 Mus musculus cDNA clone 477729.5 similar to TR:222233 222233 SUPEROXIDE DISMUTASE
05T13521	gb 037356	3.7e-70	871	Rattus sp. EST110066 Rattus sp. cDNA 3'-end
05T13521	gb 037356	6.7e-34	801	Rattus norvegicus Rattus norvegicus Rattus norvegicus mRNA, complete cds
05T13524	gb 037350	5.7e-31	831	Rattus norvegicus Rattus norvegicus Rattus norvegicus mRNA, complete cds
05T13515	gb w93148	4.0e-103	841	Mus musculus House homeobox protein (hox-1.11) gene, complete cds
05T13556	gb w00748	1.9e-129	971	Mus musculus m048109.1 Soares mouse p1NH19.5 Mus musculus cDNA clone 332667.5
05T13550	gb 103386	7.9e-132	971	Rattus norvegicus Rattus norvegicus (clone RA102-5/8) zinc finger protein mRNA, 3' end cds
05T13561	gb w13785	5.1e-64	991	Mus musculus m04921.1 Soares mouse p1NH19.5 Mus musculus cDNA clone 318256.5 similar to SW:M27.0AT
05T13567	gb AA050004	2.4e-40	701	P24051 40S K100000.1 PROTEIN S27. [11] Mus musculus m373007.1 Soares mouse embryo m0611.5 14.5 Mus musculus cDNA clone 478477.5
05T13571	gb w7236	2.4e-113	911	Mus musculus m03107.1 Soares mouse embryo m0611.5 14.5 Mus musculus cDNA clone 351188.5
05T13575	gb AA080212	6.0e-90	931	p1NH19.5 Mus musculus cDNA clone 351188.5 Mus musculus cDNA clone 351188.5 Mus musculus cDNA clone 351188.5 Mus musculus cDNA clone 351188.5
05T13579	gb w574622	1.1e-39	761	Galus gallus p1H1.121 subunit protein phosphatase 1H 21 subunit regulatory subunit (chicken, giltzard smooth muscle, mRNA, 1500 nt)
05T13582	gb w71150	1.5e-74	921	Mus musculus H. musculus 2PAC Xisterna P1000000 group A Collecting gene, exon 6
05T13601	gb w09708	1.1e-110	891	homo sapiens human gamma glutamyl exchange factor p522 mRNA, complete cds
05T13602	gb w12062	2.3e-107	901	homo sapiens YF6105.1 Homo sapiens cDNA clone 22483.5
05T13604	gb w42756	4.9e-119	841	Rattus norvegicus Rat 24-kDa subunit of mitochondrial HADH dehydrogenase mRNA, 3' end
05T13608	gb w031994	5.4e-101	871	homo sapiens human ornithine-dependent

Figure 8 cont'd.

Accession	Gene	Protein	Accession	Gene	Protein
U013609	gb AA165901	protein kinase catalytic subunit (DNA-PKcs) mRNA, complete cds Mus musculus m17500.1 r1 Soares mouse lymph node NMH19 Mus musculus cDNA clone 637010 5' similar to NMH192.6	2.4e-139	961	
U013611	gb AA028590	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 46182 5' similar to NMH192.6	2.1e-152	971	
U013612	gb AA028591	CE00805 UBC Homo sapiens m121612.1 r1 Soares mouse cDNA clone 46182 5' similar to NMH192.6	1.4e-47	711	
U013615	gb AA14951	RECEPTOR-RELATED-1 (ROR1) mRNA, complete cds Homo sapiens m121612.1 r1 Soares mouse cDNA clone 46182 5' similar to NMH192.6	1.3e-101	901	
U013617	gb AA14971	Mus musculus m121612.1 r1 Soares mouse cDNA clone 46182 5' similar to NMH192.6	1.7e-36	761	
U013621	gb AA021146	Mus musculus m121612.1 r1 Soares mouse cDNA clone 46182 5' similar to NMH192.6	1.4e-109	911	
U013622	gb AA0494	HEBRANE A1 PROTEIN (H1) Mus sp. gamma-phosphorylase kinase (alternatively spliced) [liver, muscle, brain/C. genomic, 4201 nt, segment 4 of 4]	3.1e-11	941	
U013623	gb AA0494	HEBRANE A1 PROTEIN (H1) Mus sp. gamma-phosphorylase kinase (alternatively spliced) [liver, muscle, brain/C. genomic, 4201 nt, segment 4 of 4]	3.1e-204	961	
U013629	gb AA05910	parathyroid tumor NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.0e-35	861	
U013631	gb AA05813	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 367657 5' similar to NMH19.5	7.6e-94	931	
U013634	gb AA01819	parathyroid tumor NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	5.4e-71	931	
U013700	gb AA03821	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	4.9e-171	991	
U013703	gb AA03821	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	7.0e-71	821	
U013704	gb AA048648	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	4.6e-68	931	
U013708	gb AA002215	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	7.4e-89	971	
U013716	gb AA014685	PROSPERIN (PROSPERIN) Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	0.2e-119	901	
U013729	gb AA019303	Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	2.9e-97	851	
U013731	gb AA01502	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.3e-131	931	
U013735	gb AA015275	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	5.4e-100	971	
U013757	gb AA017924	HEBRANE A1 PROTEIN (H1) Mus sp. gamma-phosphorylase kinase (alternatively spliced) [liver, muscle, brain/C. genomic, 4201 nt, segment 4 of 4]	2.6e-99	831	
U013759	gb AA04840	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	7.6e-51	971	
U013767	gb AA0536	5'-end GEM-563106 Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	5.2e-39	691	
U013775	gb AA018282	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.6e-57	971	
U013788	gb AA014126	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	9.7e-55	101	
U013792	gb AA01544	Oxidoreductase U17 SUBUNIT Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	9.5e-67	971	
U013807	gb AA02698	small subunit, complete cds Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.8e-51	801	
U013818	gb AA02438	sapiens cDNA unidentified cloning vector Mucor m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.8e-48	961	
U013819	gb AA05632	vector pLXEN, complete genome Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.8e-15	811	
U013827	gb AA04630	SP72005.10 CE00629 Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.2e-67	841	
U013831	gb AA07077	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.5e-121	991	
U013839	gb AA06008	Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.4e-103	861	
U013843	gb AA02190	hypothetical 43.5K protein Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	2.8e-51	881	
U013847	gb AA01986	180K12: UNRESOLVED Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.3e-173	941	
U013941	gb AA01037	SH3 domain protein Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.0e-135	841	
U013958	gb AA056135	11-tine-finger transcription factor (CTCF) mRNA, complete cds Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	4.7e-237	971	
U013964	gb AA019493	prothymosin alpha Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	9.8e-33	951	
U013969	gb AA01525	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	4.4e-100	851	
U013979	gb AA010405	SH3 domain protein Homo sapiens m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.8e-97	951	
U013983	gb AA029308	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.2e-108	861	
U013995	gb AA02430	Mus musculus m121612.1 r1 Soares mouse embryo NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	8.0e-102	921	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.2e-80	941	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	8.7e-81	871	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	3.9e-32	841	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.2e-88	761	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.6e-36	811	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.4e-122	971	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	2.8e-143	961	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	1.1e-88	821	
U013999	gb AA020459	placenta NMH19.5 14.5 Mus musculus cDNA clone 321363 3' similar to NMH19.5	6.6e-37	771	

Figure 8 cont'd.

OST4196	gb W41301	3.1e-39	99%	Mus musculus mc4j06.r1 Soares mouse pJMF19.5 Mus musculus cDNA clone 351323 5'.
OST4223	gb AA203787	2.7e-89	90%	Mus musculus m60f12.r1 Soares mouse lymph node NIMIN Mus musculus cDNA clone 643823 5'.
OST4228	gb S51016	9.3e-205	92%	Dos taurus E2(25K)multitubulinating enzyme [cattle, thymus, mRNA, 825 nt]
OST4229	gb 231263	4.8e-70	97%	Mus musculus H. musculus expressed sequence tag MTE577
OST4235	gb W53187	3.0e-173	97%	Mus musculus md19e07.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone J68820 5' similar to MF:C3D5.9 CE01849
OST4243	gb AA048921	2.3e-40	86%	Mus musculus m47e11.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 475276 5' similar to gb U13705 Mus musculus domesticus C318L/RJ Placenta glutathione (HOUSE)
OST4245	gb H10216	9.9e-80	75%	Homo sapiens m405.a1 Homo sapiens cDNA clone 46710 3'.
OST4247	gb AA023146	1.5e-115	96%	Mus musculus m67b03.r1 Soares mouse placenta 40Bp13.5 14.5 Mus musculus cDNA clone 455981 5' similar to SW:AP HUMAN Q04941 INTESTINAL MEMBRANE A4 PROTEIN. (1)
OST4251	gb AA070774	8.7e-154	98%	Homo sapiens m53e11.r1 Stratagene fibroblast (4937212) Homo sapiens cDNA clone 529412 3'.
OST4254	gb W54737	2.4e-82	10%	Mus musculus md10a04.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 367950 5'.
OST4258	gb AA013789	4.3e-169	90%	Mus musculus mhl3d03.r1 Soares mouse placenta 40Bp13.5 14.5 Mus musculus cDNA clone 442373 5' similar to PIR:JC2472 JC2472 RE protein - human
OST4281	gb U16175	4.0e-40	63%	Mus musculus Mus musculus thrombospondin 3 (Thbs3) gene, partial cds and mucin 1 (Muc1) gene, complete cds
OST4293	gb AA007519	8.9e-52	81%	Homo sapiens zh98e12.r1 Soares fetal liver spleen INFLS S1 Homo sapiens cDNA clone 429358 5'.
OST4298	gb AA000024	1.4e-135	96%	Mus musculus mg3e06.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 425602 5' similar to gb X01920.rna2 H. musculus CSHPX gene (HOUSE)
OST4315	gb H18210	6.4e-62	96%	Mus musculus Mouse transcription factor S-11, clone 5513
OST4319	gb J04636	2.0e-127	95%	Mus musculus Mouse glutathione S-transferase class mu (GST5-5) mRNA, complete cds
OST4391	gb W43526	9.6e-55	94%	Mus musculus mc29e04.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 354750 5'.
OST4388	gb H13524	2.6e-111	90%	Mus musculus House serum amyloid A pseudogene (psl-SAA)
OST4393	gb H16778	4.7e-45	82%	Homo sapiens Y33a08.31 Homo sapiens cDNA clone 128630 3'.
OST4002	gb AA000314	1.9e-112	96%	Mus musculus m31e07.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 425700 5'.
OST4003	gb L37297	2.9e-121	91%	Mus musculus Mus musculus (clone 86) myeloid secondary granule protein mRNA
OST4011	gb L26664	2.0e-155	94%	Mus musculus Mus musculus expressed sequence tag EST F032
OST4028	gb D87470	7.5e-93	92%	Homo sapiens human mRNA for KIAA0280 gene, partial cds
OST4033	gb AA084704	2.2e-54	88%	Homo sapiens m05c04.s1 Stratagene hurr neuron (8937233) Homo sapiens cDNA clone 546559 3' similar to TM:G00529 G00529 NADH OXIDOREDUCTASE SUBUNIT
OST4051	gb F03500	7.6e-63	86%	Homo sapiens H. sapiens partial cDNA sequence: clone c-11d08
OST4061	gb W0618	3.1e-118	97%	Mus musculus mc10h12.r1 Soares mouse pJMF19.5 Mus musculus cDNA clone J48167 5'.
OST4070	gb W16515	6.0e-135	94%	Mus musculus m67g12.r1 Soares mouse pJMF19.5 Mus musculus cDNA clone J35398 5'.
OST4073	gb X82021	2.0e-105	91%	Rattus norvegicus R. norvegicus mRNA for heat shock-related protein 1
OST4074	gb D63704	3.3e-140	86%	Rattus norvegicus Rat mRNA for dihydrodipicolinate, complete cds
OST4106	gb W5804	1.1e-84	93%	Mus musculus m67a06.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 400594 5'.
OST4114	gb W20730	6.5e-90	96%	Mus musculus m69g01.r1 Soares mouse pJMF19.5 Mus musculus cDNA clone 317296 5'.
OST4131	gb AA044274	2.4e-33	69%	Homo sapiens ak54h03.s1 Soares pregnant uterus NBPU Homo sapiens cDNA clone 486677 3'.
OST4134	gb H11489	3.0e-84	85%	Rattus sp. EST105564 Rattus sp. cDNA 3' end
OST4140	gb W1052	3.7e-121	91%	Mus musculus m62701.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone J88729 5' similar to SW:YB18 YEAST P31812 HYPOTHETICAL 13.6 KD PROTEIN IN PET112-IL51 INTENSEMIC REGION. (1)
OST4142	gb C07091	5.7e-74	89%	Rattus norvegicus similar to none
OST4144	gb X56135	4.4e-41	81%	Mus musculus House mRNA for prothymosin alpha
OST4148	gb W4510	1.5e-134	91%	Mus musculus m60b09.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 367841 5' similar to PIR:AS6059 AS6059 protein-tyrosine-phosphatase
OST4149	gb U16393	2.6e-111	96%	Mus musculus Mus musculus transcription factor TFE8 mRNA, partial cds
OST4154	gb X56046	1.3e-161	96%	Mus musculus Mouse mRNA (clone 46604-16) for hypothetical protein A
OST4155	gb X05900	3.5e-58	85%	Rattus norvegicus (pRbeta B1-1) beta1-cryodellin (pRbeta B1-1)
OST4166	gb U53859	8.0e-169	90%	Rattus norvegicus Rattus norvegicus caldesmon alpha subunit (csl1) mRNA, partial cds
OST4174	gb U11395	1.3e-38	84%	Mus musculus Mus musculus X inactive specific transcript (Xist) gene, fused M6-14A, fragment 2
OST4191	gb X61507	2.0e-75	81%	Mus musculus M. musculus HOK-3.5 clone
OST4192	gb W6357	2.2e-83	82%	Mus musculus m49h12.r1 Soares mouse embryo NBHE13.5 14.5 Mus musculus cDNA clone 408455 5' similar to SW:GLYM HUMAN P14897 SERINE HYDROXYMETHYLTRANSFERASE, MITOCHONDRIAL
OST4194	gb W14635	8.9e-38	87%	Mus musculus mc31e07.r1 Soares mouse pJMF19.5 Mus musculus cDNA clone 350148 5'.